ELSEVIER

Gene 259 (2000) 31–43

GENE
AN INTERNATIONAL JOURNAL ON GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

The compositional evolution of vertebrate genomes

Giorgio Bernardi [a,b,*]

[a] *Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Napoli 80121, Italy*
[b] *Laboratoire de Génétique Moléculaire, Institut Jacques Monod, Paris 75005, France*

Received 15 May 2000; received in revised form 24 July 2000; accepted 25 August 2000
Received by T. Gojobori

Abstract

The compositional evolution of vertebrate genomes is characterized: (i) by one predominant conservative mode, in which nucleotide changes occur, but the base composition of DNA sequences in general, and of coding sequences in particular, does not change; and (ii) by three different shifting or transitional modes, in which nucleotide changes are accompanied by changes in the base composition of sequences. 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.

This review will present first the intragenomic shifts, the 'major shifts' and the 'minor shift', and then the 'whole-genome', or 'horizontal', shift. In each case, the shifts were preceded and followed by a conservative mode of evolution. This review expands on a previous one [Bernardi, Gene 241 (2000) 3–17], and summarizes the evidence that the changes of the compositional patterns of the genome and their maintenance are controlled by Darwinian natural selection.

Keywords: Genomics; Isochores; Mutational bias; Natural selection; Random drift

1. Introduction

Many years ago, I started investigations on the organization of the nuclear genome of vertebrates and of a model system, the mitochondrial genome of yeast. In both cases, the experimental approach was the most elementary one: the analysis of the base composition of the genomes under consideration, as fractionated by density gradient centrifugation in the presence of sequence-specific ligands. This compositional approach led to new insights into the organization of the eukaryotic genome. Later, when comparative studies were made on the organization of the nuclear genomes of vertebrates, compositional differences were shown to have an evolutionary relevance. Indeed, investigating the compositional evolution of the vertebrate genome led to a better understanding of the role of natural selection in molecular evolution, an important topic discussed in a recent article (Bernardi, 2000), and expanded here. 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. Fig. 1 presents a scheme of the conservative mode of evolution and of the three kinds of compositional shifts that we 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.

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, fractionation of a nuclear mammalian DNA (the bovine DNA) using Cs_2SO_4 density gradient centrifugation in the presence of

Abbreviations: GC, molar ratio of guanine+cytosine in DNA; GC_3, GC of third codon positions.

* Tel.: +39-081-5833-300. fax: +39-081-7641-355.
E-mail address: bernardi@alpha.szn.it (G. Bernardi)

0378-1119/00/$ - see front matter
PII: S0378-1119(00)00441-8

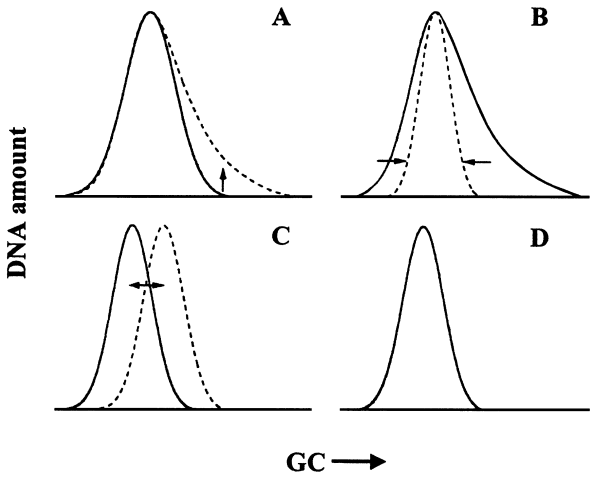

Fig. 1. The 'major' (**A**), the 'minor' (**B**) and the 'whole-genome' or 'horizontal' (**C**) compositional transitions (shifts) in the genomes of vertebrates. This scheme displays CsCl profiles and the changes undergone as a consequence of the compositional shift under consideration. CsCl profiles 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. (Modified from Bernardi et al., 1997.)

sequence-specific DNA ligands, such as Ag^+ (Corneo et al., 1968) or BAMD, 3,6-bis(acetato mercuri methyl)1,4-dioxane (Cortadas et al., 1979), revealed (Filipski et al., 1973) a remarkable compositional heterogeneity of the 'main band DNA', namely of the nuclear DNA except for satellite and ribosomal DNAs. Subsequent investigations (Thiery et al., 1976; Macaya et al., 1976; Cortadas et al., 1979; Olofsson and Bernardi, 1983; Bernardi et al., 1985) demonstrated that DNAs from all warm-blooded vertebrates exhibit high compositional heterogeneities and strongly asymmetrical CsCl bands, whereas DNAs from cold-blooded vertebrates are generally characterized by low compositional heterogeneities and by only slightly asymmetrical CsCl bands (see Fig. 1A). Both differences, in heterogeneity and asymmetry, later studied in more detail (Cuny et al., 1981; Bernardi and Bernardi, 1990a,b), are due to the presence in the genomes of warm-blooded vertebrates of a small percentage (about 15%) of GC-rich DNA molecules (GC is the molar ratio of guanine+cytosine in DNA) 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 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 (Bernardi et al., 1985, 1988; Perrin and Bernardi, 1987; Mouchiroud et al., 1987, 1988; Bernardi and Bernardi, 1991). Indeed, coding sequences from cold-blooded vertebrates are compositionally relatively homogeneous and generally characterized 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* (as initially shown by Perrin and Bernardi, 1987; Mouchiroud et al., 1987, 1988; Bernardi and Bernardi, 1991). 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 average, compared with 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* (Fig. 2).

Two additional observations were made concerning the compositional transitions under consideration. First, codon frequencies and codon usage were essentially unchanged for the orthologous genes that had not been affected by the major shift, whereas they were drastically changed for those that had been affected (Cruveiller

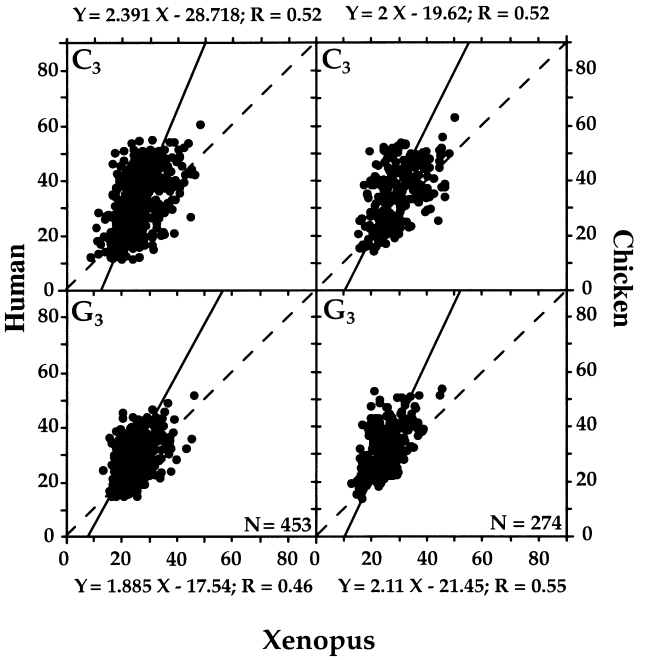

Fig. 2. Correlation between G_3 and C_3 values of orthologous genes from human, or chicken, and *Xenopus*. The orthogonal regression lines are shown together with the diagonals (dashed lines). The equations of the regression lines and the correlation coefficients are indicated. *N* is the number of gene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

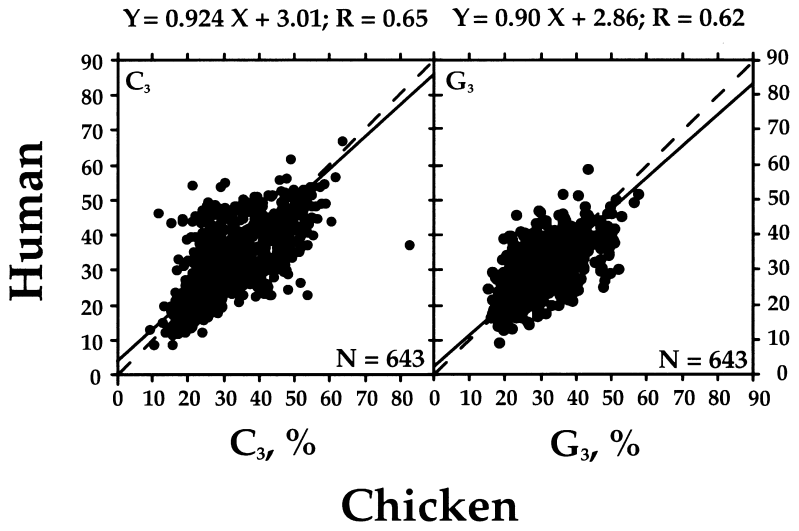

Fig. 3. Correlation between G_3 and C_3 values of orthologous genes of human and chicken. Other indications are as in Fig. 2.

et al., 2000). Second, the regression lines between GC_3, G_3 and C_3 values of orthologous genes from human and chicken showed a high correlation coefficient and coincided with the diagonal, indicating that the genes that had not undergone the transition and those that had were, to a large extent, the same sets of genes in the two species (Fig. 3). Since very good correlations exist between GC_3 and GC levels of extended DNA regions flanking the coding sequences (Bernardi et al., 1985; Clay et al., 1996), the results of Fig. 3 also indicate that large regions surrounding orthologous genes underwent (or, alternatively, did not undergo) the compositional transition and account for the parallelism between compositional patterns seen at the DNA level and at the coding sequence level (see Bernardi, 1995). This observation stresses the link between the major shifts as observed at the DNA level and at the coding sequence level.

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. 4). 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 (Sabeur et al., 1993; Mouchiroud and Bernardi, 1993). In Fig. 4, the regression lines of the human/calf plots pass through the origin and are characterized 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 between different avian orders (Kadi et al., 1993; Mouchiroud and Bernardi, 1993). Since mammalian orders diverged some 100 Myr 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 (see Bernardi, 2000). (ii) They occurred (and were similar) in the independent ancestral lines of mammals and birds (exceptions will be mentioned later). (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. 4, 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 that time 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 50% 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.

Two additional points to be made here are the following. First, codon frequencies and codon usage were essentially unchanged for the human/*Xenopus* or

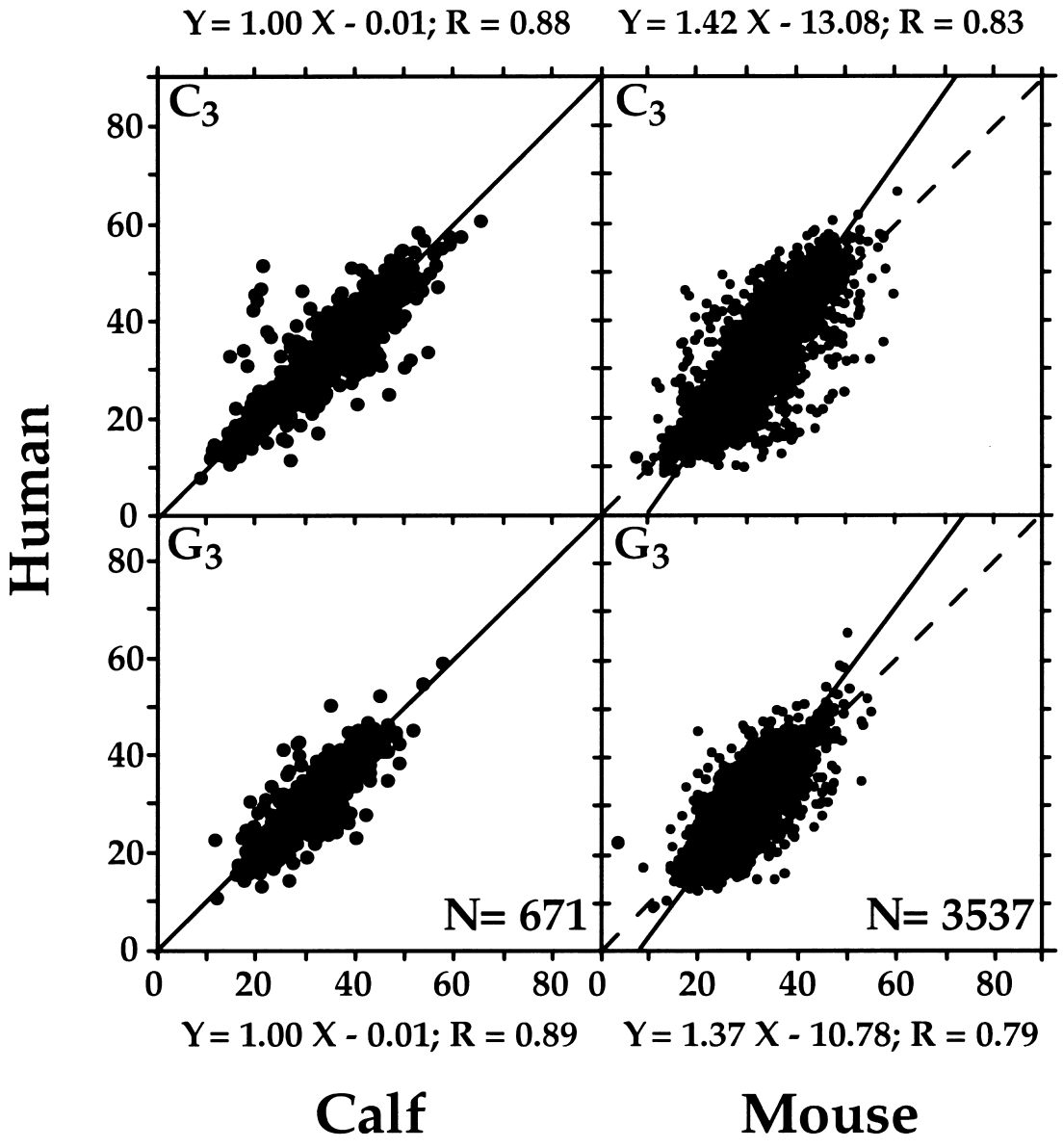

Fig. 4. Correlation between G_3 and C_3 values of orthologous genes from human and calf or mouse. Other indications are as in Fig. 2.

chicken/*Xenopus* orthologous genes that had not been affected by the major shift, whereas they were drastically changed for those that had been affected (Fig. 5). Interestingly, no significant difference was found for either GC-poor or GC-rich orthologous genes form human and chicken (Cruveiller et al., 2000). Second, compositional transitions 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 with fishes, or amphibians (Jabbari et al., 1997); in fact, these changes apparently occurred already between amphibians and reptiles (which seem to exhibit the lower methylation pattern of warm-blooded vertebrates); (ii) CpG islands were formed (Bernardi, 1989; Aïssani and Bernardi, 1991a,b; CpG islands are regulatory sequences about 1 kb in size, located 5′ of GC-rich genes, and characterized by high levels of GC and unmethylated CpG doublets); (iii) changes took place in the AUG initiator context of GC-rich human genes relative to genes from cold-blooded vertebrates (Pesole et al., 1999); (iv) T bands appeared in metaphase chromosomes; (v) karyotype changes and speciation increased (Bernardi, 1993a). Interestingly, changes (ii) and (iii) indicate that the major shifts also altered regulatory mechanisms.

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 (Bernardi and Bernardi, 1986) was that they were due to natural selection, namely "the preservation of favourable varia-

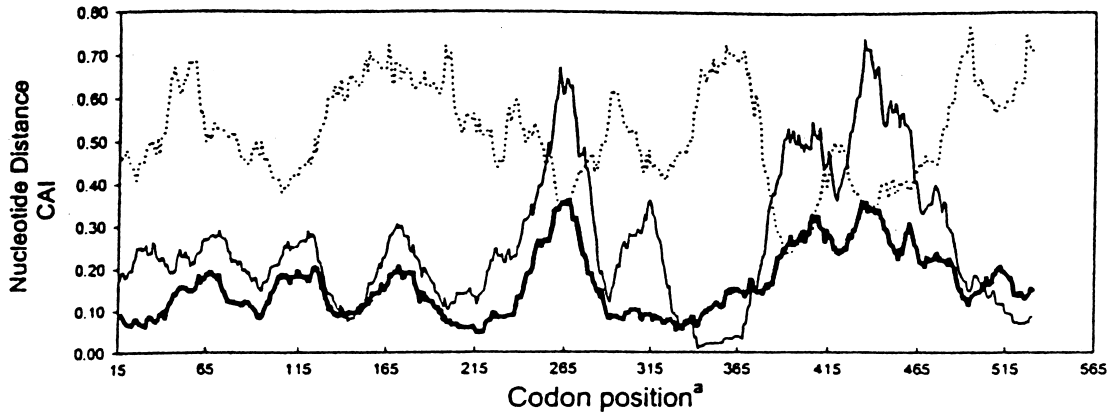

Fig. 5. Profiles of synonymous distance (thin line), non-synonymous distance (thick line) and Codon Adaptation Index (dotted line) for the GP63 gene of *Leishmania*. The window size used was 30 codons, shifting one codon at a time. Each window was labelled according to the codon falling in the middle, so that the first window is assigned to codon 15. The correlation coefficients were calculated using non-overlapping windows to insure the independence of sampling points. (From Alvarez-Valin et al., 2000a.)

tions and the rejection of injurious variations'' (Darwin, 1859); 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 (Saccone et al., 1992, 1993, 1996) that largely coincide with the T(elomeric) bands previously identified as particularly resistant to thermal denaturation (Dutrillaux, 1973). As for RNA, abundant evidence indicates that high GC levels stabilize RNA structures (see, for example, Hasegawa et al., 1979; Wada and Suyama, 1986; Galtier and Lobry, 1997).

As for protein stability, three lines of evidence support it. First of all, an increase in GC_3 of coding sequences corresponds to an increase in GC_2 and GC_1, as shown by the existence of a positive universal compositional correlation among the three codon positions (D'Onofrio and Bernardi, 1992). It also corresponds to an increase in quartet codons, to a decrease in duet codons, to an increase in certain amino acids (alanine, valine), to a decrease in other amino acids (lysine) and to an overall increase in the hydrophobicity of the encoded proteins (D'Onofrio et al., 1999). 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* (Cruveiller et al., 1999), indicating that the compositional genome transition between cold- and warm-blooded vertebrates was accompanied by changes in structural features of the encoded proteins that stabilize them. Incidentally, very recent work has shown that the strongest correlation of hydrophobicity is not so much with GC_3 as with C_3 (Jabbari and Bernardi, in preparation).

That the hydrophobicity changes are functionally significant is indicated by the fact that the systematically higher hydrophobicity of prokaryotic relative to eukaryotic proteins is accompanied by a cysteine level in prokaryotic proteins half as high as that found in vertebrates (D'Onofrio et al., 1999). In turn, this suggests that the stabilizing higher hydrophobicity of prokaryotic proteins was replaced in eukaryotic proteins by a higher number of disulphide bridges. Since the difference in hydrophobicity of prokaryotic and eukaryotic proteins has the same magnitude as that between proteins encoded by GC-rich and GC-poor genes in the human genome, if the former is functionally significant, as suggested by the compensatory changes in disulphide bridges, so should be the latter. In other words, the higher hydrophobicity of proteins encoded by genes that underwent the compositional transition should be functionally relevant. Very recent observations on paralogous proteins encoded by genes characterized by different GC levels (Rayko et al., in preparation) provide an additional check of this point.

Another finding in favor of the functional significance of the changes is that regions of coding sequences corresponding to α helix, β sheet and coil structures in the proteins are characterized by different levels of individual nucleotides in all codon positions and by different substitution rates (Chiusano et al., 1999, 2000). 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 α helix compared with regions corresponding to β sheet.

Very recently, a sliding window analysis performed on 19 *Leishmania* genes encoding the surface metalloproteinase GP63, whose crystallographic structure is known, showed: (i) that the rate of synonymous substitutions along the gene is highly correlated with both the

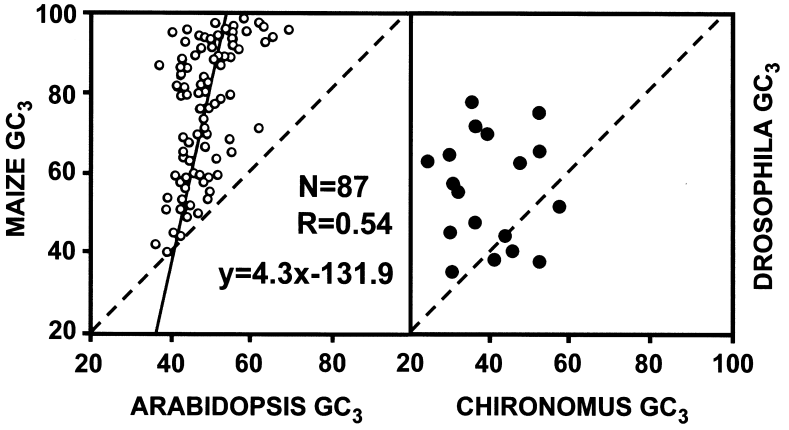

Fig. 6. (Left) GC_3 values of maize genes are plotted against the GC_3 values of their homologs from *Arabidopsis* genes (Carels et al., 1998). (Right) GC_3 values of *Drosophila* genes are plotted against the GC_3 values of their homologs from *Chironomus*.

rate of amino acid substitution and codon usage; and (ii) that there is a clear relationship between the rates and the tertiary structure of the encoded proteins, since all divergent segments are located on the surface of the molecule and facing one side (almost parallel to the cellular membrane) on the exposed surface of the organism (Alvarez-Valin et al., 2000; see Fig. 5).

Another line of (admittedly circumstantial) evidence in favor of natural selection is provided by observations made in other organisms which are rather closely related to each other, yet phylogenetically distant from mammals and birds. Indeed, most genes from *Gramineae* exhibit a much higher GC_3 level relative to orthologous genes from dicots (Salinas et al., 1986; Matassi et al., 1999; Carels et al., 1998; see Fig. 6, left). It should be noted that, while this compositional transition occurred between an ancestral monocot and *Gramineae*, it can only be investigated at present by comparing orthologous genes from *Gramineae* and dicots (which are ancestral to monocots) due to the lack of appropriate samples of genes from monocots ancestral to *Gramineae*. Likewise, genes from *Drosophila melanogaster* and *Anopheles gambiae*, which exhibit a wide compositional range (with the GC-rich regions enriched in genes; Jabbari and Bernardi, 2000; Myers et al., 2000), were systematically GC-richer (Fig. 6, right) than their orthologs from the GC-poorer, compositionally homogeneous, genome of *Chironomus thummi* (Jabbari and Bernardi, in preparation).

Both the situation found in *Gramineae* relative to dicots and that of *Drosophila* relative to *Chironomus* are very reminiscent of what was described before for the human/*Xenopus* or chicken/*Xenopus* comparisons. These four intragenomic compositional transitions may, therefore, either be similar by sheer coincidence, or because they are due to similar factors. We argued that *Gramineae*, in contrast to the reference dicots, originated from hot, arid regions and had to stand very high maximal temperatures for geological times (Bernardi et al., 1988). Likewise, *Drosophila* larvae are exposed to temperatures as high as 45°C (Feder, 1996), and *Anopheles* is a tropical insect, whereas *Chironomus thummi* is a dipteran from sub-arctic regions. It is, therefore, possible that a common factor, selection for an increased thermal stability, is responsible for these similar compositional genome transitions.

At this point, it is difficult to draw any conclusion about the causes of the transition in methylation and CpG levels of vertebrates which seems to have already occurred at the time of the appearance of reptiles. We speculated (Jabbari et al., 1997), however, that the lower 5 mC level of warm-blooded vertebrates may be due to a higher deamination rate related to their higher body temperature. Indeed, the deamination of 5 mC residues in double-stranded DNA has a strong temperature dependence (Shen et al., 1994). The fact that the genomes of reptiles are apparently similar to those of warm-blooded vertebrates in their methylation might be accounted for by the high body temperature reached by many reptiles. It should be mentioned that the deamination (C→T) rate is much higher than the mutation rate at CpG doublets and that the latter is largely dependent upon the efficiency of the mismatch repair mechanism (Shen et al., 1994; Yang et al., 1996). This mechanism might, however, be overwhelmed by the five-fold increased production of mismatches due to the increase in temperature from approximately 20 to 37°C. Under this hypothesis, a methylation decrease would not require the permanent higher body temperature of warm-blooded vertebrates to become effective, whereas other properties, such as the formation of GC-rich isochores, CpG islands, and T bands, might require it.

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 trans-

ition while others did not might also be explained by 'regional' mutational biases (other explanations have been discussed and ruled out elsewhere; see Bernardi, 2000). It was pointed out (Bernardi et al., 1988), 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 (Sueoka, 1988). Alternatively, repair could be more efficient in some regions, for instance in transcribed sequences (Bohr et al., 1985; see also Balajee and Bohr, 2000 for a recent review). We might even strengthen here these additional hypotheses by suggesting that regional biases might concern gene-poor compared with gene-rich regions, and also that they could be related to the different replication timings of gene-poor and gene-rich regions (Federico et al., 1998, 2000).

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 are also characterized 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 (Matassi et al., 1999; Montero et al., 1990; Carels et al., 1998), and with different secondary structures of proteins (Chiusano et al., 1999, 2000); (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 unrelated to vertebrates; (v) why G and R bands (De Sario et al., 1996, 1997), replication banding (Federico et al., 1998, 2000; Saccone et al., 1999), and synonymous rate regions (Matassi et al., 1999), 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; and, most importantly, (vii) why the mutational bias GC→AT that was found by analyzing mutational substitution matrices of both GC-poor and GC-rich human genes (F. Alvarez-Valin et al.) did not lead to a GC-poor compositional pattern in the human genome; and why very similar differences in C_3 compared to G_3 were found when comparing codon frequencies of orthologous genes of human and chicken (Cruvellier et al., 2000). Incidentally, the finding that some genes underwent the major shift, whereas others did not, may be related to the stringency of structural and functional requirements of the encoded proteins. This possibility seems to be supported by the abundance in GC-rich isochores of housekeeping genes, whose protein products are indispensable for cell life, and by the abundance of tissue-specific genes in GC-poor isochores (see Bernardi, 2000).

3. The minor shift

Although characteristic differences between the CsCl profiles of murids and other mammals were first observed by Thiery et al. (1976), only more detailed investigations performed later definitely proved that the mouse genome did not exhibit the GC-richest DNA components that were present in the human genome (Zerial et al., 1986; Salinas et al., 1986). These findings were confirmed by investigations at the coding sequence level, which demonstrated in murid genes characterized by extreme base compositions a *minor, intragenomic, shift* (see Fig. 1B) relative to orthologous genes from mammals exhibiting the general pattern (Mouchiroud et al., 1988). Indeed, 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. 4). Now, it is known that murids exhibit: (i) rates of synonymous substitutions that are higher, by a factor of five to 10, relative to human coding sequences (Wu and Li, 1985; Gu and Li, 1992); (ii) a defective repair system (see Holliday, 1995); and (iii) a compositional pattern that is derived from the general mammalian pattern (Galtier and Mouchiroud, 1998). 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 previous section). It is also remarkable that this also leads to the 'randomization' of non-coding sequences, causing, for instance, an 'erosion' of CpG islands (Aïssani and Bernardi, 1991a,b; Matsuo et al., 1993). In the case of the α 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

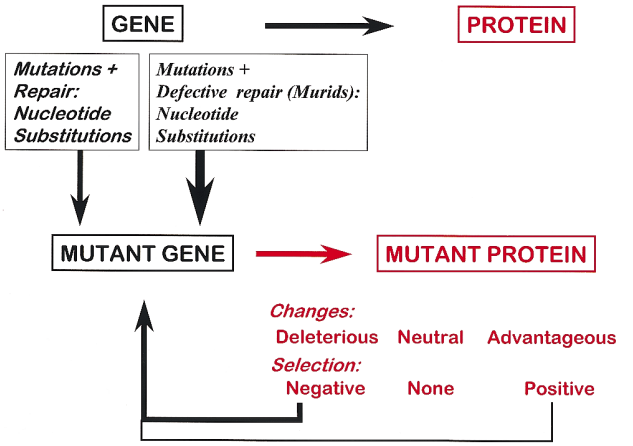

Fig. 7. Scheme of the equilibrium between mutational input (nucleotide substitutions) and negative selection. The equilibrium on the left is that of genes from mammals showing the general mammalian pattern, the equilibrium on the right is the case of genes from murids. (Modified from Bernardi, 2000.)

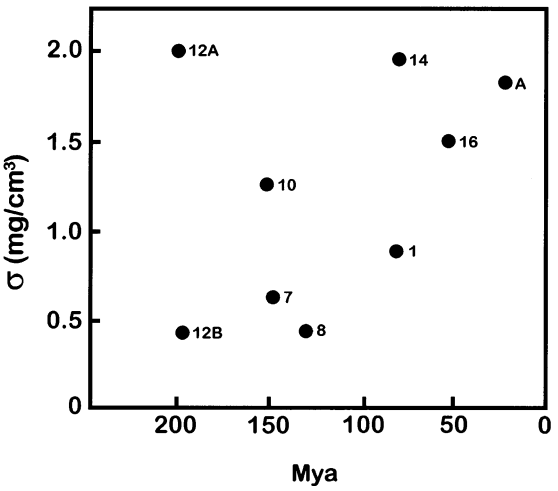

Fig. 8. Standard deviations of average modal buoyant densities in CsCl, σ, of DNAs from fish species belonging to the same family, or order, are plotted against the time of the appearance of these groups. Groups are: Lamniformes, 1; Cypriniformes, 7; Salmonidae, 8; Gadiformes, 10; Aplocheilidae, 12A; Cyprinodontidae, 12B; *Aphyosemion*, A; Perciformes, 14; Tetraodontiformes, 16. Numbers correspond to those of table 2 of Bernardi and Bernardi (1990b) from which this figure was taken.

was reached (see Fig. 7), as witnessed by the very similar compositional pattern of different murids (Douady et al., 2000a,b), 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. This is the approximate time of divergence for rodents that have a defective repair system and higher rates of nucleotide substitutions (like murids) compared with other rodents (like Histricomorphs and Sciuromorphs) that have a stringent repair system and correspondingly share the general mammalian pattern. Needless to say, the minor shift has nothing to do with body temperature.

4. The whole-genome shifts

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 with amphibians and reptiles (Bernardi and Bernardi, 1990a,b). 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 characterized 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 (Fig. 8); and (iii) they are much more frequent and much larger than the horizontal shifts found in mammalian and avian orders (Bernardi and Bernardi, 1990a,b; see Table 1 and Fig. 9).

Among the three points mentioned above, the first one indicates an essential 'randomness' in the primary events responsible for the changes and is compatible with variations in their directionality (AT→GC; GC→AT). The second point indicates, in addition, that there is no cumulative effect, in that the genomes of ancient orders do not show more spreading of their average composition compared with those of recent ones, again in line with the randomness and also with possible variations in the directionality of changes. The third point suggests that homeostasis leads to more stable compositional patterns of the genome (see Table 1); in other words, negative selection appears to

Table 1
Average modal buoyant densities of DNAs from vertebrates[a]

	Number of species	ρ_0 (g/cm³)	$\sigma(\rho_0)$ (mg/cm³)
Chondrichthyes	12	1.7035	1.1
Osteichthyes	110	1.7014	2.2
Amphibians	5	1.7020	1.7
Reptiles	13	1.7019	1.8
Birds[b]	8	1.7000	1.0
Mammals[c]	10	1.7000	0.7

[a] From Bernardi and Bernardi (1990b). Modal buoyant density is linearly related to average GC level. 1 mg/cm³ corresponds approximately to 1% GC.
[b] From Kadi et al. (1993).
[c] Only Eutherians were considered.

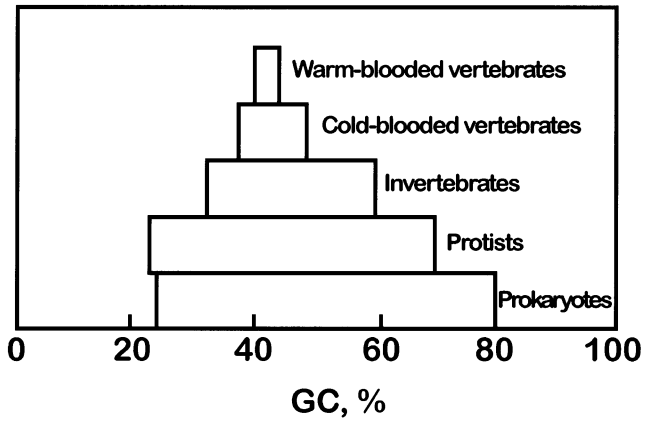

Fig. 9. GC level ranges of DNAs from warm- and cold-blooded vertebrates, invertebrates, protists, and prokaryotes. (From Bernardi and Bernardi, 1990b.)

restrict the range of tolerated mutations in warm-blooded vertebrates, that are more controlled in their environment, compared with cold-blooded vertebrates, which use more diverse ecological niches (see below).

4.1. The horizontal shifts 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 compositions of prokaryotic genomes is that genome compositions shift because of directional mutations due to biases in replication enzymes (Freese, 1962; Sueoka, 1962). 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 (Bernardi et al., 1988). Indeed, 'mutator mutations' may lead not only to highly increased mutation rates, but also to strong biases in base substitutions. In the laboratory, such biases have only been detected in mutational hot spots (Nghiem et al., 1988; Wu et al., 1990). Overall changes in genome composition such as those claimed by Cox and Yanofsky (1967), although still accepted by many authors, fall, in fact, within the experimental error (Bernardi, 1993b). 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, 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. The first one 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. The 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. 9). Since the potentially relevant mutations in the replication machinery responsible for different mutational biases are, presumably, 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. 9 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, it is possible that these two kinds of mutations overlap; indeed, whole-genome shifts have been observed in mammalian genomes which had undergone either major or minor shifts (see Table 1); on the other hand, some cold-blooded vertebrates may show changes comparable to major shifts; for instance, the crocodile and the turtle that were studied by Aïssani and Bernardi (1991a,b) showed this phenomenon. (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 may specifically act on the mutations in the genes encoding the replication machinery.

5. 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 with 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, the previous section and below). Interestingly, compositional changes also affected regulatory sequences and, therefore, gene expression. This is a whole, important area that deserves further work.

The explanation originally proposed for the major shifts (Bernardi and Bernardi, 1986), 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 α helix, β 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*, from *Chironomus* and *Drosophila*/*Anopheles*. In each case, the genome that underwent the transition experienced over a long period a higher body temperature compared with the one that did not. It is not impossible that the genomes of some reptiles (such as the crocodile and the turtle investigated by Aïssani and Bernardi (1991a,b; see previous section) may have undergone similar compositional changes. In fact, if this point could be firmly established, it would provide an 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 with 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 (Alvarez-Valin, personal communication) reinforces the need for natural selection to maintain the mammalian compositional pattern. This need seems to be less strong for the GC-poor isochores (see also Zoubak et al., 1995).

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'.

6. Some general remarks

Some final remarks deserve to be made at this point.

(i) Natural selection, as discussed in this review in connection with the major shifts, is essentially a negative (or stabilizing) selection. This can easily be 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.

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 for 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 neighboring 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 (see Rynditch et al., 1998 for a review).

(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 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 vertebrates may cope with higher environmental temperatures over relatively short times with 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 horizontal shifts of prokaryotes and unicellular eukaryotes and even the narrower ones of vertebrate genomes are also related to temperature, as unduly extrapolated by several authors from the conclusions drawn for the major shift of vertebrates (see Hughes, 1999 for a recent example), 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 initiated in our laboratory a long time ago 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 centered 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.

Acknowledgements

Thanks are due to Fernando Alvarez, Giacomo Bernardi, Giuseppe D'Onofrio, Hector Musto and, especially, Oliver Clay for critical reading and discussions.

References

Aïssani, B., Bernardi, G., 1991a. CpG islands: features and distribution in the genome of vertebrates. Gene 106, 173–183.

Aïssani, B., Bernardi, G., 1991b. CpG islands, genes and isochores in the genome of vertebrates. Gene 106, 185–195.

Alvarez-Valin, F., Tort, J.F., Bernardi, G., 2000a. Non-random spatial distribution of synonymous substitutions in the Leishmania GP63 gene. Genetics. in press.

Alvarez-Valin, F., et al., 2000b. Gene submitted.

Balajee, A.S., Bohr, V.A., 2000. Genomic heterogeneity of nucleotide excision repair. Gene 250, 15–30.

Bernardi, G., 1989. The isochore organization of the human genome. Annu. Rev. Genet. 23, 637–661.

Bernardi, G., 1993a. Genome organization and species formation in vertebrates. J. Mol. Evol. 37, 331–337.

Bernardi, G., 1993b. The vertebrate genome: isochores and evolution. Mol. Biol. Evol. 10, 186–204.

Bernardi, G., 1995. The human genome: organization and evolutionary history. Annu. Rev. Genet. 29, 445–476.

Bernardi, G., 2000. Isochores and the evolutionary genomics of vertebrates. Gene 241, 3–17.

Bernardi, G., Bernardi, G., 1986. Compositional constraints and genome evolution. J. Mol. Evol. 24, 1–11.

Bernardi, G., Bernardi, G., 1990a. Compositional patterns in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 265–281.

Bernardi, G., Bernardi, G., 1990b. Compositional transitions in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. 31, 282–293.

Bernardi, G., Bernardi, G., 1991. Compositional properties of nuclear genes from cold-blooded vertebrates. J. Mol. Evol. 33, 57–67.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M., Rodier, F., 1985. The mosaic genome of warm-blooded vertebrates. Science 228, 953–958.

Bernardi, G., Mouchiroud, D., Gautier, C., Bernardi, G., 1988. Compositional patterns in vertebrate genomes: conservation and change in evolution. J. Mol. Evol. 28, 7–18.

Bernardi, G., Hughes, S., Mouchiroud, D., 1997. The major compositional transitions in the vertebrate genome. J. Mol. Evol. 44, S44–S51.

Bohr, V.A., Smith, C.A., Okumoto, D.S., Hanawalt, P.C., 1985. DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. Cell 40, 359–369.

Carels, N., Hatey, P., Jabbari, K., Bernardi, G., 1998. Compositional properties of homologous coding sequences from plants. J. Mol. Evol. 46, 45–53.

Chiusano, M.L., D'Onofrio, G., Alvarez-Valin, F., Jabbari, K., Colonna, G., Bernardi, G., 1999. Correlations of nucleotide substitution rates and base composition of mammalian coding sequences with protein structure. Gene 238, 23–31.

Chiusano, M.L., Alvarez-Valin, F., Di Giulio, M., D'Onofrio, G., Ammirato, G., Colonna, G., Bernardi, G., 2000. Second codon positions of genes and the secondary structures of proteins: relationships and implications for the origin of the genetic code. Gene. in press.

Clay, O., Cacció, S., Zoubak, S., Mouchiroud, D., Bernardi, G., 1996. Human coding and non-coding DNA: compositional correlations. Mol. Phylogen. Evol. 5, 2–12.

Corneo, G., Ginelli, E., Soave, C., Bernardi, G., 1968. Isolation and characterization of mouse and guinea pig satellite DNAs. Biochemistry 7, 4373–4379.

Cortadas, J., Olofsson, B., Meunier-Rotival, M., Macaya, G., Bernardi, G., 1979. The DNA components of the chicken genome. Eur. J. Biochem. 99, 179–186.

Cox, E.C., Yanofsky, C., 1967. Altered base ratios in the DNA of an *Escherichia coli* mutator strain. Proc. Natl. Acad. Sci. USA 88, 1895–1902.

Cruveiller, S., Jabbari, K., D'Onofrio, G., Bernardi, G., 1999. Different hydrophobicities of orthologous proteins from *Xenopus* and man. Gene 238, 15–21.

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

Cuny, G., Soriano, P., Macaya, G., Bernardi, G., 1981. The major components of the mouse and human genomes: preparation, basic properties and compositional heterogeneity. Eur. J. Biochem. 111, 227–233.

Darwin, C., 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. Murray, London.

De Sario, A., Geigl, E.-M., Palmieri, G., D'Urso, M., Bernardi, G., 1996. A compositional map of human chromosome band Xq28. Proc. Natl. Acad. Sci. USA 93, 1298–1302.

De Sario, A., Roizés, G., Allegre, N., Bernardi, G., 1997. A compositional map of the cen-q21 region of human chromosome 21. Gene 194, 107–113.

D'Onofrio, G., Bernardi, G., 1992. A universal compositional correlation among codon positions. Gene 110, 81–88.

D'Onofrio, G., Jabbari, K., Musto, H., Bernardi, G., 1999. The correlation of protein hydropathy with the composition of coding sequences. Gene 238, 3–14.

Douady, C., Carels, N., Clay, O., Catzeflis, F., Bernardi, G., 2000a. Diversity and phylogenetic implications of CsCl profiles from rodent DNAs. Mol. Phylogen. Evol. in press.

Douady, C., et al., 2000b. Mol. Phylogen. Evol. submitted.

Dutrillaux, B., 1973. Nouveau système de marquage chromosomique: les bandes T. Chromosoma 41, 395–402.

Feder, M.E., 1996. Ecological stress and evolutionary physiology of the stress proteins and the stress response: the *Drosophila melanogaster* model in animals and temperature. In: Johnston, I.A., Bennett, A.F. (Eds.), Phenotypic and Evolutionary Adaptation. Cambridge University Press, Cambridge.

Federico, C., Saccone, S., Bernardi, G., 1998. The gene-richest bands of human chromosomes replicate at the onset of the S-phase. Cytogenet. Cell Genet. 80, 83–88.

Federico, C., Andreozzi, L., Saccone, S., Bernardi, G., 2000. Gene density in the GIEMSA bands from human chromosome. Chromosome Res. in press.

Filipski, J., Thiery, J.P., Bernardi, G., 1973. An analysis of the bovine genome by Cs_2SO_4–Ag^+ density gradient centrifugation. J. Mol. Biol. 80, 177–197.

Freese, E., 1962. On the evolution of base composition of DNA. J. Theor. Biol. 3, 82–101.

Galtier, N., Lobry, J.R., 1997. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. 44, 632–636.

Galtier, N., Mouchiroud, D., 1998. Isochore evolution in mammals: a human-like ancestral structure. Genetics 150, 1577–1584.

Gu, X., Li, W.H., 1992. Higher rates of amino acid substitution in rodents than in humans. Mol. Phylogen. Evol. 1, 211–214.

Hasegawa, S., Yasunaga, T., Miyata, T., 1979. Secondary structure of MS2 phage RNA and bias in code word usage. Nucleic Acids Res. 7, 2073–2079.

Holliday, R., 1995. Understanding Ageing. Cambridge University Press, Cambridge.

Hughes, A.L., 1999. Adaptive Evolution of Genes and Genomes. Oxford University Press, Oxford.

Jabbari, K., Bernardi, G., 2000. The distribution of genes in the *Drosophila* genome. Gene 247, 287–292.

Jabbari, K., Cacció, S., Païs de Barros, J.P., Desgrés, J., Bernardi, G., 1997. Evolutionary changes in CpG and methylation levels in the genome of vertebrates. Gene 205, 109–118.

Kadi, F., Mouchiroud, D., Sabeur, G., Bernardi, G., 1993. The compositional patterns of the avian genomes and their evolutionary implications. J. Mol. Evol. 37, 544–551.

Macaya, G., Thiery, J.P., Bernardi, G., 1976. An approach to the organization of eukaryotic genomes at a macromolecular level. J. Mol. Biol. 108, 237–254.

Matassi, G., Sharp, P.M., Gautier, C., 1999. Chromosomal location effects or gene sequence evolution in mammals. Curr. Biol. 9, 786–791.

Matsuo, K., Clay, O., Takahashi, T., Silke, J., Schaffner, W., 1993. Evidence for erosion of mouse CpG islands during mammalian evolution. Som. Cell Mol. Gen. 6, 543–555.

Montero, L.M., Salinas, J., Matassi, G., Bernardi, G., 1990. Gene distribution and isochore organization in the nuclear genome of plants. Nucleic Acids Res. 18, 1859–1867.

Mouchiroud, D., Bernardi, G., 1993. Compositional properties of coding sequences and mammalian phylogeny. J. Mol. Evol. 37, 109–116.

Mouchiroud, D., Fichant, G., Bernardi, G., 1987. Compositional compartmentalization and gene composition in the genome of vertebrates. J. Mol. Evol. 26, 198–204.

Mouchiroud, D., Gautier, C., Bernardi, G., 1988. The compositional distribution of coding sequences and DNA molecules in humans and murids. J. Mol. Evol. 27, 311–320.

Myers, E.W, et al., 2000. A whole-genome assembly of *Drosophila*. Science 287, 2196–2204.

Nghiem, Y., Cabrera, M., Cupples, C.G., Miller, J.H., 1988. The *mut Y* gene: a mutator locus in *Escherichia coli* that generates CC–TA transversions. Proc. Natl. Acad. Sci. USA 85, 2709–2713.

Olofsson, B., Bernardi, G., 1983. Organization of nucleotide sequences in the chicken genome. Eur. J. Biochem. 130, 241–245.

Perrin, P., Bernardi, G., 1987. Directional fixation of mutations in vertebrate evolution. J. Mol. Evol. 26, 301–310.

Pesole, G., Bernardi, G., Saccone, C., 1999. Isochore specificity of AUG initiator context of human genes. FEBS Lett. 464, 60–62.

Rynditch, A.V., Zoubak, S., Tsyba, L., Tryapitsina-Guley, N., Bernardi, G., 1998. The regional integration of retroviral sequences into the mosaic genomes of mammals. Gene 222, 1–16.

Sabeur, G., Macaya, G., Kadi, F., Bernardi, G., 1993. The isochore patterns of mammalian genomes and their phylogenetic implications. J. Mol. Evol. 37, 93–108.

Saccone, S., De Sario, A., Della Valle, G., Bernardi, G., 1992. The highest gene concentrations in the human genome are in T bands of metaphase chromosomes. Proc. Natl. Acad. Sci. USA 89, 4913–4917.

Saccone, C., De Sario, A., Wiegant, J., Rap, A.K., Della Valle, G., Bernardi, G., 1993. Correlations between isochores and chromosomal bands in the human genome. Proc. Natl. Acad. Sci. USA 90, 11 929–11 933.

Saccone, S., Cacció, S., Kusuda, J., Andreozzi, L., Bernardi, G., 1996. Identification of the gene-richest bands in human chromosomes. Gene 174, 85–94.

Saccone, S., Federico, C., Solovei, I., Croquette, M.F., Della Valle, G., Bernardi, G., 1999. Identification of the gene-richest bands in human prometaphase chromosomes. Chrom. Res. 7, 379–386.

Salinas, J., Zerial, M., Filipski, J., Bernardi, G., 1986. Gene distribution and nucleotide sequence organization in the mouse genome. Eur. J. Biochem. 160, 469–478.

Shen, J.C., Rideout, W.M., Jones, P.A., 1994. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res. 22, 972–976.

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.

Thiery, J.P., Macaya, G., Bernardi, G., 1976. An analysis of eukaryotic

genomes by density gradient centrifugation. J. Mol. Biol. 108, 219–235.

Wada, A., Suyama, A., 1986. Local stability of DNA and RNA secondary structure and its relation to biological function. Prog. Biophys. Mol. Biol. 47, 113–157.

Wu, C.L., Li, W.H., 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82, 1741–1745.

Wu, T.-H., Clarke, C.H., Marinus, M.G., 1990. Specificity of *Escherichia coli mutD* and *mutL* mutator strains. Gene 87, 1–5.

Yang, A.S., Gonzalgo, M.L., Zingg, J.M., Millar, R.P., Buckley, J.D., Jones, P.A., 1996. The rate of CpG mutation in Alu repetitive elements within the p53 tumor suppressor gene in the primate germline. J. Mol. Biol. 258, 240–250.

Zerial, M., Salinas, J., Filipski, J., Bernardi, G., 1986. Gene distribution and nucleotide sequence organization in the human genome. Eur. J. Biochem. 160, 479–485.

Zoubak, S., D'Onofrio, G., Cacció, S., Bernardi, G., Bernardi, G., 1995. Specific compositional patterns of synonymous positions in homologous mammalian genes. J. Mol. Evol. 40, 293–307.