J Mol Evol (1997) 44(Suppl 1):S44–S51

JOURNAL OF MOLECULAR EVOLUTION

The Major Compositional Transitions in the Vertebrate Genome

Giorgio Bernardi,[1] Sandrine Hughes,[2] Dominique Mouchiroud[2]

[1] Laboratoire de Genetique Moleculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France
[2] Laboratoire de Biometrie, Genetique et Biologie des Populations, U.M.R. 5558, Universite Claude Bernard, 69600 Villeurbanne, France

Received: 26 August 1996 / Accepted: 4 September 1996

Abstract. The vertebrate genome underwent two major compositional transitions, between therapsids and mammals and between dinosaurs and birds. These transitions concerned a sizable part (roughly one-third) of the genome, the gene-richest part of it, and consisted in an increase in GC levels (GC is the molar fraction of guanine + cytosine in DNA) which affected both coding sequences (especially third codon positions) and noncoding sequences. These major transitions were studied here by comparing GC_3 levels (GC_3 is the GC of third codon positions) of orthologous genes from *Xenopus*, chicken, calf, and man.

Key words: Base composition — Evolution — Isochores — Orthologous mammalian genes

Introduction

Twenty years ago, it[1] was discovered that vertebrate genomes (1) are mosaics of isochores, long DNA segments which show a remarkably uniform base composition, and can be partitioned into a small number of families characterized by different GC levels (Macaya et al. 1976; GC is the molar fraction of guanine + cytosine in DNA); and (2) exhibit different compositional patterns in cold- and warm-blooded vertebrates (Thiery et al. 1976). These discoveries, which concerned ''main-band'' DNA (and not satellite or minor DNA components, such as ribosomal DNA), raised crucial questions about the compositional evolution of eukaryotic genomes and its cause(s) (see Bernardi and Bernardi 1986; Bernardi et al. 1988; Bernardi 1989, 1993a,b, 1995), a problem previously investigated in prokaryotes (Freese, 1962; Sueoka, 1962).

The initial observations of Thiery et al. (1976) were followed by detailed studies (Hudson et al. 1980; Pizon et al. 1984; Bernardi and Bernardi 1990 a,b) which showed that the genomes of cold-blooded vertebrates differ from those of warm-blooded vertebrates in exhibiting a much smaller compositional heterogeneity. As far as mammals are concerned, it may be recalled here that while about two-thirds of the human genome (which is representative of most mammalian genomes; see below) are made up of GC-poor, or L(ight) isochores, which are compositionally close to the genomes of most cold-blooded vertebrates, the remaining third is formed by three GC-rich, or H(eavy), isochore families, H1, H2, and H3, the last of which reaches GC levels that are not attained in the genomes of cold-blooded vertebrates (Bernardi et al. 1985; Zerial et al. 1986). These results indicate that the genomes of cold-blooded vertebrates and mammals are separated by a ''major compositional transition'' in which a sizable part (roughly one-third) of the vertebrate genome underwent a GC increase. In avian genomes, the GC increase attained even higher values than in mammalian genomes (Cortadas et al. 1979; Olofsson and Bernardi 1983).

This conclusion was also reached when coding sequences were investigated. Indeed, the distributions of GC_3 values (the GC levels of third codon positions) of *Xenopus* and human genes showed differences similar to those seen at the level of DNA fragments in that in *Xenopus* the GC_3 distribution was narrower and did not reach the high values shown by mammalian coding se-

Correspondence to: G. Bernardi

quence or the higher one exhibited by avian coding sequences (Mouchiroud et al. 1987; Bernardi et al. 1988; Bernardi 1993a,b, 1995).

Another approach developed in order to study the major transitions was to compare the GC_3 values of orthologous genes. When this approach was used, most genes tested exhibited higher GC_3 values in mammals and birds than in cold-blooded vertebrates (Perrin and Bernardi 1987; Bernardi et al. 1988; Bernardi and Bernardi 1991; Kadi et al. 1993).

Apart from the two ''major compositional transitions'' separating cold-blooded vertebrates from mammals and birds, some ''minor transitions'' were detected, for instance, between murids and most other mammals. These will not be examined in detail here (see, however, the Discussion).

An important step forward in understanding the major compositional transitions was made when information was obtained on the gene distribution in the vertebrate genomes. While the earliest experiments on gene localization already showed a strikingly non-uniform gene distribution in the human genome, with gene concentration increasing from GC-poor to GC-rich isochores (Bernardi et al. 1985), subsequent, more detailed investigations provided precise estimates of gene concentrations in different isochore families. These cover at least a 17-fold range, with relative gene concentrations ranging from 55.5% in H3, to 32.2% in H2, 9.0% in H1, and 3.3% in L isochores (Mouchiroud et al. 1991; Zoubak et al. 1996).

On the other hand, recent information on the gene distribution of cold-blooded vertebrates has supported the early suggestion (Bernardi and Bernardi 1990b) that all vertebrate genomes must share a similar, non-uniform gene distribution. This suggestion was based on the general consideration that, if such was not the case, the transitions from cold- to warm-blooded vertebrates should have been accompanied by a most unlikely massive reshuffling of genes. Indeed, if GC-rich genes, or genes having undergone GC increases, are visualized as originally scattered all over the genome of cold-blooded vertebrates, they would have needed to cluster in the GC-rich isochores of warm-blooded vertebrates.

At present, the idea of a non-uniform, and similar, gene distribution in all vertebrate genomes is supported by at least three lines of evidence. The first one is that the GC-richest human isochores of the H3 family hybridize, under conditions in which repeated sequences cannot do so, not only with the GC-richest isochores of other mammals and of birds (Caccio et al. 1994) but also with the GC-richest isochores of cold-blooded vertebrates (Perani et al., paper in preparation). The second line of evidence is the existence of syntenic regions between the genomes of vertebrates, even when they are as widely separated as mammals and fishes (see Walter et al. 1993; Elgar 1996). The third line of evidence comes from the significant correlation found between GC_3 values of orthologous coding sequences from man (or other mammals) and chicken, or man and *Xenopus* (Bernardi and Bernardi 1991; Kadi et al. 1993; Bernardi 1993b, 1995; and this paper).

Here, we have used the most recent releases of the DNA sequences of orthologous genes in the data base from human and chicken and human and *Xenopus* in order to reexamine the major compositional transitions of vertebrate genomes as studied at the third codon positions. Although the human/calf data used here come from a previous release, recent additions have not changed the GC_3 plot or the conclusions. The results shown extend previous studies (Perrin and Bernardi 1987; Bernardi et al. 1988; Bernardi and Bernardi 1991; Kadi et al. 1993; Bernardi 1993b, 1995) and lead to a better understanding of the major compositional transitions in vertebrates.

Materials and Methods

Sequences from GenBank (release 95) were managed using the HOVERGEN software (Duret et al. 1994), which is based on the ACNUC system (Gouy et al. 1985) and provides a very efficient way to select homologous genes from vertebrates and to test if they are orthologous. The GC_3 frequencies studied correspond to the percentage of GC in the third positions of aligned coding sequences. For *Xenopus*, the orthologous genes were distributed in three equal classes according to their GC_3 frequencies: a GC_3-poor class (<47.4%); a GC_3-intermediate class (comprised between 47.4% and 55.7%), and a GC_3-rich class (>55.7%).

Results

The main features of the human/*Xenopus* GC_3 plot involving 209 gene pairs (see Fig. 1A) are the following: (1) the correlation coefficient, 0.57, is highly significant ($P < 0.0001$) and (2) the slope of the orthogonal regression line is 2.47.

In the human/chicken GC_3 plot, which concerned 361 gene pairs (Fig. 1B), (1) the correlation coefficient, 0.69, is highly significant ($P < 0.0001$); (2) the slope, 0.9, is close to unity; and (3) the spread of points increases toward higher GC values.

The human/calf GC_3 plot involving 245 gene pairs (Fig. 1C) shows (1) an extremely high correlation coefficient, 0.97, and (2) a slope barely different from unity, 0.97.

Figure 2A shows the correlation between GC_3 values of orthologous genes from human and chicken which are shared by *Xenopus.* In this particular human/chicken plot concerning 94 gene pairs, the correlation coefficient, 0.75, is slightly higher than in the case of Fig. 1B, whereas the slope, 0.89, is practically the same, and an increasing spread of higher GC_3 values is also observed. The point of interest is that 75% of GC-rich genes of

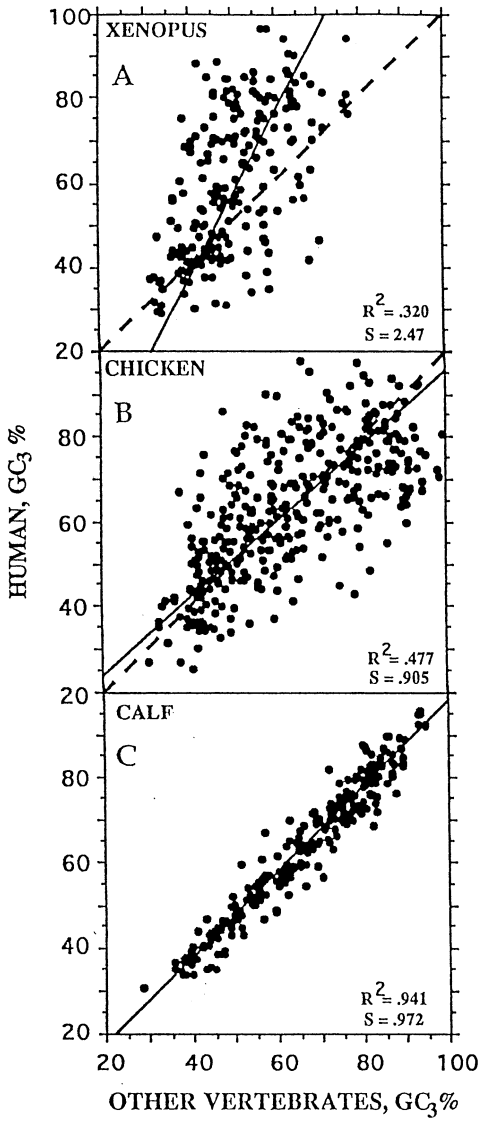

Fig. 1. Correlation between GC_3 value of orthologous genes from **(A)** human and *Xenopus* ($N = 209$), **(B)** human and chicken ($N = 361$), and **(C)** human and calf ($N = 245$). *N* is the number of gene pairs analyzed. The orthogonal regression lines are shown together with the diagonal (broken) lines.

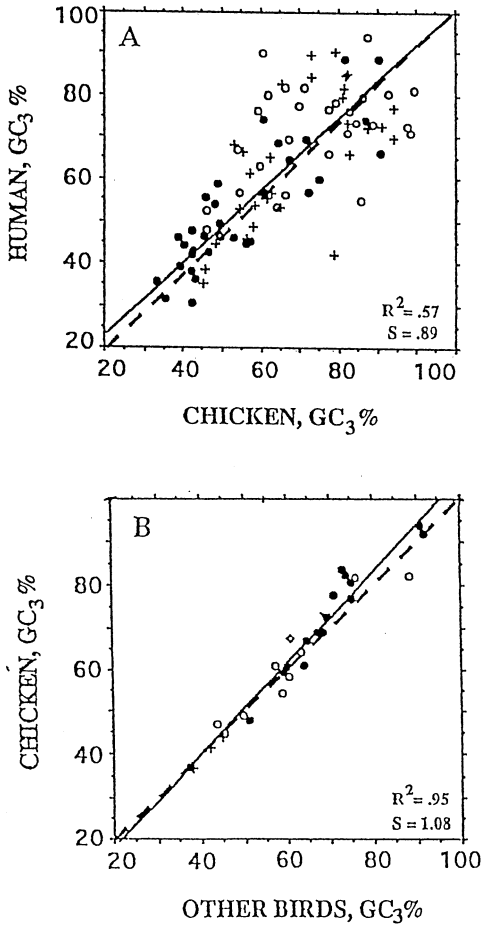

Fig. 2. Correlation between GC_3 values of orthologous genes **(A)** shared by human, chicken, and *Xenopus* ($N = 94$). ● $GC_3 < 47.4\%$; + $47.4\% < GC_3 < 55.7\%$; o $GC_3 > 55.7\%$ in *Xenopus*. **B** From chicken and other birds (● Galliformes; ○ Anseriformes; + Columbiformes; ◇ Passeriformes). Broken lines indicate diagonals.

Xenopus ($GC_3 > 55.7$) have homologs in human and/or chicken characterized by GC_3 values higher than 60%, and, conversely, 71% of GC-poor genes of *Xenopus* ($GC_3 < 47.4\%$) have homologs in human and/or chicken with GC_3 values lower than 60%.

As in the case of the mammalian (human/calf) GC_3 plot, and in spite of the long divergence time, the chicken/other-birds plot (Fig. 2B) also shows (1) an extremely high correlation coefficient, 0.98, and (2) a slope close to unity, 1.08.

Discussion

A first preliminary remark is that the present investigation is strictly focused on the two major compositional transitions of vertebrate genomes. It is known that ''minor'' transitions also occurred both in mammals (Salinas et al. 1986; Mouchiroud et al. 1987; Sabeur et al. 1993; Mouchiroud and Bernardi 1993) and in fishes (Bernardi and Bernardi 1990b). Such transitions did not lead, however, to the generation of GC-rich isochores in only a part of the genome similar to those observed when comparing cold-blooded vertebrates with mammals and birds (see Fig. 3A). In the case of murids, the transition relative to most other mammals, in fact, corresponds to a change in the width of the compositional distribution of both isochores and genes (see Fig. 3B); in the case of fishes the transition corresponds to overall changes in composition (see Fig. 3C). These ''minor'' transitions will be discussed elsewhere, because they correspond to phenomena fundamentally different from those responsible for the major transitions.

Another preliminary remark is that the basic features of compositional plots concerning third codon positions have not significantly changed over the past few years, in spite of a very large increase in the gene samples available. This suggests that, although minor details may still

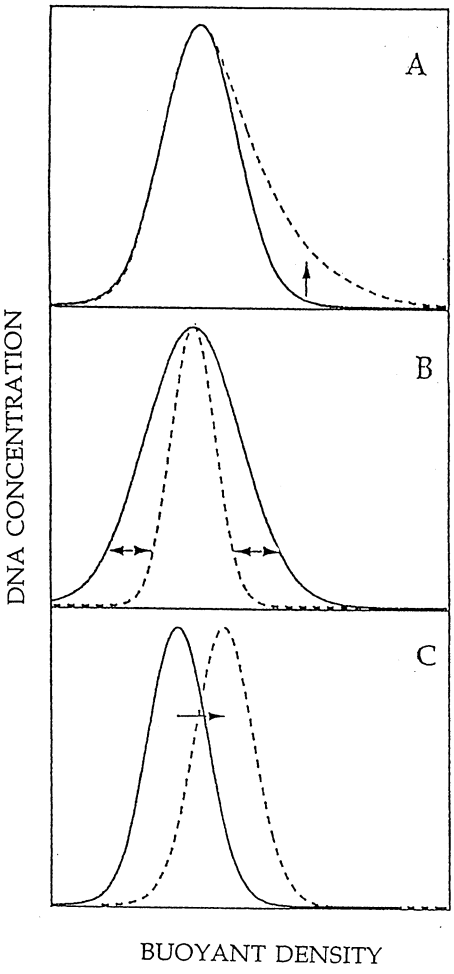

Fig. 3. The ''major'' **(A)** and ''minor'' **(B, C)** compositional transitions in the genomes of vertebrates; **B** refers to the transition between the general mammalian pattern and the murid pattern, **C** refers to the transitions among cold-blooded vertebrates. The scheme presented refers to DNA.

vary as the number of available orthologous sequences increases, no important change in the conclusions drawn here should be expected.

The Human/Xenopus Plot

The major cold- to warm-blooded compositional genome transitions can be studied at the level of GC_3 values only in the case of *Xenopus* (Fig. 1A) because data for other cold-blooded vertebrates are still scarce. It should be stressed, however, that a plot of the data available for *Xenopus*/other *Anura* (not shown) indicates a compositional similarity among these amphibians.

The high correlation coefficient of the human/*Xenopus* plot indicates that, in spite of the very large phylogenetic distance separating man from *Xenopus*, GC_3 values of GC-poor genes from the two species are still close to each other, on the average, whereas the GC_3 divergence increases with increasing GC_3. Moreover, the correlation indicates that in *Xenopus* the GC-richest genes are those that have undergone the largest increase in GC at the transition. In other words, while the GC-poor genes underwent compositional changes showing no directionality on average, the GC-richer genes underwent directional changes which were increasingly larger at increasing GC levels. While the former results fit with a compositionally conservative mode of evolution, the latter ones indicate a transitional mode, in which nucleotide substitutions are accompanied by compositional changes (Bernardi et al. 1988). Since a correlation exists between GC_3 and the GC level of the isochores in which genes are embedded (Aïssani et al. 1991; Clay et al. 1996), one can distinguish in warm-blooded vertebrates a paleogenome characterized by the absence of directional change, from a neogenome, in which directional changes took place (Bernardi 1989).

The scatter of the points in the diagram region corresponding to compositional changes may be due, at least in part, to the amplification of gene families between amphibians and mammals, to their subsequent scatter through the genome, and to their present location in different isochores of the mammalian genome (as well as to some accidental paralogous, instead of orthologous, comparisons). The distribution of genes from the actin and vimentin families in different isochores of the genomes of warm-blooded vertebrates is well documented (Dodemont et al. 1982). Another well-known case is that of α- and β-globin genes which are still linked together in two *Xenopus* loci whereas they are separated in mammals, where the β-globin gene cluster has remained GC-poor while the α-globin gene cluster has become very GC-rich (Bernardi et al. 1985).

At the same time, the results of Fig. 1A strongly support the idea of a generally similar gene distribution in *Xenopus* and human, in that genes located in GC-poor and GC-rich isochores in human also are located in GC-poor and GC-rich isochores in *Xenopus*, respectively, even if GC levels are much lower in the latter case than in the former one (see also Fig. 2A, which indicates that the distributions of GC_3 values from orthologous genes are similar in Xenopus, chicken, and man).

The Human/Chicken Plot

The human/chicken GC_3 plot (Fig. 1B) can be considered representative of a mammalian/avian plot because most mammals (those showing the general pattern; Sabeur et al. 1993; Mouchiroud and Bernardi 1993) are characterized by a GC_3 correlation similar to that shown by human and calf (Fig. 1C) and because the compositional properties of avian DNAs (Kadi et al. 1993) and the chicken/other-birds plot (Fig. 2B) do not suggest any compositional transition in birds. The human/chicken GC_3 plot is of special interest because these two lines of vertebrates diverged, independently of each other (see below), from two different reptilian lines, therapsids in the case of mammals, dinosaurs in the case of birds. Moreover, the two divergences took place 220 million

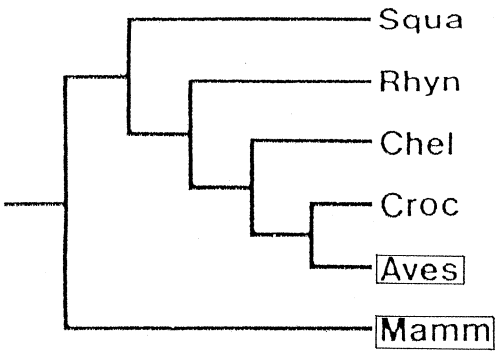

Fig. 4. Phylogenetic classification of Amniota. *Chel*, Chelonia; *Croc*, Crocodylia; *Mamm*, Mammalia; *Rhyn*, Rhyncocephalia; *Squa*, Squamata (from Lövtrup 1977). The relative order of branching off of Chelonia relative to Lepidosauria (Rhyncocephalia and Squamata) is not settled (see Caspers et al. 1996).

years ago in the case of mammals and 150 million years ago in the case of birds, the common ancestor of mammals and birds having lived 250–300 million years ago (Carroll 1987).

It should be stressed that mammals and birds do not share a common reptilian ancestor showing the same compositional pattern, because their common ancestor also gave rise to all extant reptilian orders (Fig. 4), which do not show the compositional pattern of warm-blooded vertebrates (Bernardi and Bernardi 1990a,b). The viewpoint that mammals represent the sister group of all other extant amniotes is not only supported by paleontological data (see, for instance, Carroll 1987) but also by very recent molecular data (Caspers et al. 1996). As mentioned by Caspers et al. (1996), this view goes back to Haeckel (1866), whereas the opposite view of a sister-group relation between mammals and birds, reviving the clade *Haematothermia* of Owen (1866), is not supported by paleontological or molecular data (see Hedges and Maxson 1996).

It should also be stressed that we do not know when the compositional transitions took place or how long they lasted. Indeed, transitions may have taken a very long time, possibly up to 50 Myr or so.

The significant correlation of the human/chicken plot indicates therefore a convergent (independent) compositional evolution. In other words, on average, largely the ''same'' genes underwent approximately the ''same'' compositional changes in two genome transitions which were totally independent from each other (in keeping with a similar gene distribution in all vertebrates), while no other reptilian line or any other cold-blooded vertebrate showed any similar changes (Bernardi and Bernardi 1990a,b). This means that the ''compositional order'' of mammalian and avian genes, namely, the GC rank of homologous genes, is largely the same, even in third codon positions. Because of the linear correlations holding between GC_3 and GC levels of the isochores hosting the corresponding genes (see Bernardi et al. 1985; Aissani et al. 1991; Clay et al. 1996), the ''compositional ranking order'' of the isochores is also largely the same. The above statements should be considered a valid generalization to which exceptions exist, however. To quote just one example, both α- and β-globin gene clusters of chicken are in GC-rich isochores, in contrast to those of man (see above; incidentally, in both mammals and birds the two clusters are located on different chromosomes).

The Human/Calf Plot

The human/calf plot (Fig. 1C) indicates that a compositional conservation holds for the third codon positions of all orthologous sequences in these two species, independently from their GC_3 levels, which cover a very wide range (from less than 40% to more than 90%). While the human/calf plot is shown here as an example because the largest number of orthologous sequences was available for this pair of species, plots for other mammalian pairs, whose genomes follow the general pattern (Sabeur et al. 1993), exhibit extremely similar features (not shown). As already mentioned, we are not discussing here the so-called minor transitions which separate, for example, murids from other mammals.

Although supported by a lower gene sample, an analogous conclusion on compositional conservation can be drawn from comparisons involving chicken, on the one hand, and birds belonging to a number of different orders on the other (Fig. 2B; see also Kadi et al. 1993).

If one takes into account the star phylogeny of mammalian orders, one is led to the conclusion that the compositional patterns revealed by the human/calf plot (as well as by plots involving other mammalian orders not shown here) should have also been present in their common ancestor over 65 million years ago. This justifies the concept of the (compositionally) conservative mode of evolution (Bernardi et al., 1988) which is exemplified by most mammals. If one takes into account the wide range of GC_3 values, which extend beyond 90%, it is also difficult to escape the conclusion that a compositional equilibrium was maintained in most mammalian genomes for more than 65 million years.

A very similar reasoning can also be applied to birds.

Conclusions

Figure 5 summarizes the compositional evolution of the vertebrate genome or, more precisely, of the amniote genome. Schematically, two situations were found: (1) The genes exhibiting low GC_3 values (assigned in the diagram a <50% value) basically did not undergo, on the average, any significant compositional transition, in that these genes maintained in mammals the low GC levels that they had in amphibians and in their reptilian ancestor; (2) the genes showing high GC_3 values (assigned

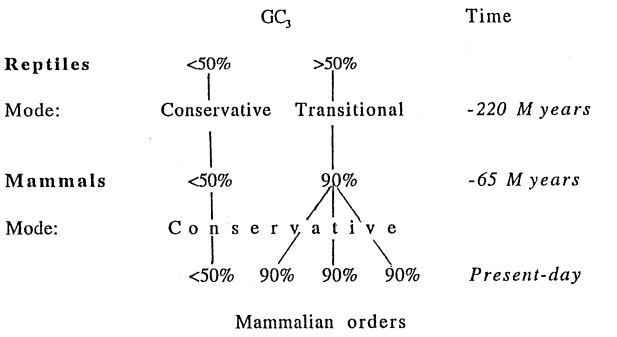

Fig. 5. Scheme of the compositional genome evolution of vertebrates (see text).

>50% values) underwent a compositional transition (arbitrarily indicated as reaching 90% GC_3 in Fig. 5). In most mammalian orders, the high GC_3 values did not change during over 65 million years of evolution. It should be noted that a scheme similar to that of Fig. 5 also applies to birds.

These situations imply that the compositional transitions affecting a minority of the isochores forming the vertebrate genome (the gene-richest isochores) have superimposed themselves onto a conserved gene distribution pattern. It should be stressed that genome size changes have obviously not affected gene distribution and compositional patterns very much. This can be easily understood in the case of polyploidization. This is, however, also the case when contraction/expansion phenomena are responsible for genome size changes, because these phenomena basically only affect intergenic sequences and introns. In fact, because these noncoding sequences are GC-poorer than coding sequences, contraction and expansion have led to GC increases or decreases, respectively, as indicated by investigations of fish genomes (Bernardi and Bernardi 1990a; Bernardi 1995).

Another feature which has been conserved throughout the evolution of the vertebrate genome is the replication pattern. This is biphasic in all vertebrate genomes, with an early and late timing (see Bernardi 1989). It is very likely that yet another feature is conserved through the evolution of the vertebrate genome, namely, the ''open'' or ''closed'' chromatin structures, because these are associated with high and low gene concentrations, respectively (see Gardiner 1995).

Several phenomena have accompanied the major transitions of vertebrates:

1. Formation of T(elomeric) bands. This seems to be the direct consequence, at the chromosomal level, of the formation of very GC-rich isochores of warm-blooded vertebrates. Indeed, T bands comprise the majority of H3 and H2 isochores, with small amounts of H1 and L isochores (Saccone et al. 1992, 1993, 1996).
2. Formation of CpG islands. CpG islands, 1-kb or so, sequences characterized by high GC levels, abundant HpaII sites (CCGG) and G/C boxes, which are regularly associated with all housekeeping genes and 40% of tissue-specific genes of warm-blooded vertebrates, appear to have arisen as a consequence of the formation of GC-rich isochores (see Bernardi 1995) since they do not exist in cold-blooded vertebrates (Bernardi 1989; Aïssani and Bernardi 1991a,b; Cross et al. 1991).
3. Species formation. Mammals exhibit a species formation rate which is, on the average, 5 times higher than cold-blooded vertebrates, and a karyotypic change rate which is close to 10 times that of cold-blooded vertebrates (Bush et al. 1977). These two phenomena were interpreted as due to social structuring and to the propensity of forming small demes (Bush et al. 1977). An alternative explanation (Bernardi 1989, 1993), not exclusive of the former, was, however, that the increased rates of species formation and karyotypic changes of warm-blooded vertebrates are associated with the formation of GC-rich isochores, which had led to the formation in warm-blooded vertebrates of T bands (see above), where recombination is most active.

Three different explanations have been proposed for the major compositional genome transitions of vertebrates. The first explanation (Bernardi and Bernardi 1986; Bernardi et al. 1988; Bernardi 1993a) was that selection eliminated the changes toward GC decreases while keeping those toward GC increases. This phenomenon affected the gene-richest regions of the ancestral genomes of mammals and birds. This explanation seems to be supported by an analysis of the frequencies and of compositional patterns of synonymous substitutions in quartet codons from orthologous mammalian genes (Caccio et al 1995; Zoubak et al. 1995).

The second explanation (Sueoka 1988, 1992, 1993) was a revival of the mutational bias hypothesis first proposed to account for the compositional changes which affected bacterial genomes (Freese 1962; Sueoka 1962). According to this explanation, biased nucleotide substitutions are the consequence of mutations in the enzymes involved in DNA replication/repair. This explanation, as applied to the major compositional transitions of vertebrates, has, however, two problems. It does not account (1) for the fact that an overall bias in replication and/or repair is not seen in all regions of the genomes under consideration, but only in gene-rich regions, or (2) for the fact that the bias is not observed in any vertebrate line other than those leading to mammals and birds, in spite of the very similar genome organization to all vertebrates (gene distribution, replication timing, and, in all likelihood, chromatin structure; see above). A solution for the first problem is the proposal (Sueoka 1988) that the different chromatin structures of gene-poor and gene-rich

regions affect the bias. The second problem remains, however, intractable.

A third explanation (Ohta 1996) reconciles the two viewpoints and, at least apparently, solves the last problem by proposing, under the assumption of the selective advantages represented by GC-rich isochores, that the evolutionary modifications in the genes of replication and/or repair enzymes are associated with such selective advantage.

These three viewpoints are just mentioned here. Indeed, the complexity of the underlying problems requires a detailed discussion which will be presented elsewhere.

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

Aïssani B, D'Onofrio G, Mouchiroud D, Gardiner K, Gautier C, Bernardi G (1991) The compositional properties of human genes. J Mol Evol 32:497–503

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

Bernardi G (1993a) The vertebrate genome: isochores and evolution. Mol Biol Evol 10:186–204

Bernardi G (1993b) The human genome organization and its evolutionary history: a review. Gene 135:57–66

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

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

Bush GL, Case SM, Wilson AC, Patton JL (1977) Rapid speciation and chromosomal evolution in mammals. Proc Natl Acad Sci USA 74:3942–3946

Caccio S, Perani P, Saccone S, Kadi F, Bernardi G (1994) Single-copy sequence homology among the GC-richest isochores of the genomes from warm-blooded vertebrates. J Mol Evol 39:331–339

Caccio S, Zoubak S, D'Onofrio G, Bernardi G (1995) Nonrandom frequency patterns of synonymous substitutions of homologous mammalian genes. J Mol Evol 40:280–292

Carroll RL (1987) Vertebrate paleontology and evolution. WH Freeman, New York

Caspers GJ, Reinders GJ, Leunissen J, Wattel J, de Jong W (1996) Protein sequences indicate that turtles branched off from the amniote tree after mammals. J Mol Evol 42:580–586

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

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

Cross S, Kovarik P, Schmidtke J, Bird A (1991) Non-methylated islands in fish genomes are GC-poor. Nucleic Acids Res 19:1469–1474

Dodemont H-J, Soriano P, Quax WJ, Ramaekers F, Lenstra JA, Groenen MAM, Bernardi G, Bloemendal H (1982) The genes coding for the cytoskeletal proteins actin and vimentin in warm-blooded vertebrates. EMBO J 1:987–994

Duret L, Mouchiroud D, Gouy M (1994) HOVERGEN: homologous vertebrate genes data base. Nucleic Acids Res 22:2360–2363

Elgar G (1996) Quality not quantity: the pufferfish genome. Hum Mol Genet 5:1437–1442

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

Gardiner K (1995) Human genome organization. Curr Opin Genet Dev 5:315–322

Gouy M, Gautier L, Attimonelli M, Lanave C, di Paola G (1985) ACNUC—a portable retrieval system for nucleic acid sequence database: logical and physical design and usage. Comput Appl Biosci 1:167–172

Haeckel E (1866) Generelle Morphologie der Organismen. G. Reiner, Berlin

Hedges SB, Maxson LR (1996) Molecular and morphology in amniote phylogeny. Mol Phylogenet Evol 6:311–314

Hudson AP, Cuny G, Cortadas J, Haschemeyer AEV, Bernardi G (1980) An analysis of fish genomes by density gradient centrifugation. Eur J Biochem 112:203–210

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

Lövtrup S (1977) The phylogeny of vertebrates. Wiley, London

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

Mouchiroud D, Bernardi B (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, D'Onofrio G, Aïssani B, Macaya G, Gautier C, Bernardi G (1991) The distribution of genes in the human genome. Gene 100:181–187

Ohta T (1996) The current significance and standing of neutral and nearly neutral theories. Bioessays 18:673–677

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

Owen R (1866) On the anatomy of vertebrates, vol 1. Longmans, London

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

Pizon V, Cuny G, Bernardi G (1984) Nucleotide sequence organization in the very small genome of a tetraodontid fish, *Arothron diadematus.* Evol J Biochem 140:25–30

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

Saccone S, De Sario A, Wiegant J, Raap AK, Della Valle G, Bernardi G (1993) Correlation between isochores and chromosomal bands in the human genome. Proc Natl Acad Sci USA 90:11929–11933

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

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

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

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 85:2563–2567

Sueoka N (1992) Directional mutation pressure, selective constraints, and genetic equilibria. J Mol Evol 34:95–114

Sueoka N (1993) Directional mutation pressure, mutator mutations, and dynamics of molecular evolution. J Mol Evol 37:137–153

Thiery JP, Macaya G, Bernardi G (1976) An analysis of eukaryotic genomes by density gradient centrifugation. J Mol Biol 108:219–235

Walter RB, Robig RL, Kozak KA, McEntire B, Morizot DC, Nairn RS (1993) Cloning and gene map assignment of the *Xiphophorus* DNA ligase 1 gene. Mol Biol Evol 10:1227–1238

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, Caccio S, Bernardi G, Bernardi G (1995) Specific compositional patterns of synonymous positions in homologous mammalian genes. J Mol Biol 40:293–307

Zoubak S, Clay O, Bernardi G (1996) The gene distribution of the human genome. Gene 174:95–102