ELSEVIER

Gene 241 (2000) 3–17

GENE
AN INTERNATIONAL JOURNAL ON GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

Review

Isochores and the evolutionary genomics of vertebrates

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 3 November 1999

Abstract

The nuclear genomes of vertebrates are mosaics of isochores, very long stretches (≫300 kb) of DNA that are homogeneous in base composition and are compositionally correlated with the coding sequences that they embed. Isochores can be partitioned in a small number of families that cover a range of GC levels (GC is the molar ratio of guanine + cytosine in DNA), which is narrow in cold-blooded vertebrates, but broad in warm-blooded vertebrates. This difference is essentially due to the fact that the GC-richest 10–15% of the genomes of the ancestors of mammals and birds underwent two independent compositional transitions characterized by strong increases in GC levels. The similarity of isochore patterns across mammalian orders, on the one hand, and across avian orders, on the other, indicates that these higher GC levels were then maintained, at least since the appearance of ancestors of warm-blooded vertebrates. After a brief review of our current knowledge on the organization of the vertebrate genome, evidence will be presented here in favor of the idea that the generation and maintenance of the GC-richest isochores in the genomes of warm-blooded vertebrates were due to natural selection.

1. Introduction

In this review, I will concentrate on investigations from our laboratory. I will first present a summary of our current knowledge on the sequence organization of the human genome (which is typical of most mammalian genomes, and shares its basic properties with avian genomes), and of the genomes of cold-blooded vertebrates. I will then describe the compositional transitions which occurred when warm-blooded vertebrates emerged from reptiles and the maintenance of the new compositional patterns in mammals and birds, respectively. Finally I will discuss the general implications of these results.

This order of presentation, which reflects the chronological development of our work, is also a logical one. Indeed, the present organization of the human genome is the result of a long evolutionary process, which has taken close to 500 million years since the earliest vertebrates. Understanding this genome organization provides the best starting point for asking precise questions about its evolutionary origin.

2. Sequence organization of the mammalian genome

The experimental approach that we followed was based on the study of the most elementary property of the genome, its nucleotide composition, more precisely the frequencies of nucleotides in DNA molecules. This approach (reviewed by Bernardi et al., 1973), the only one that was possible before DNA sequencing became available, still is extremely useful. Indeed, it could be, and was, very easily moved from DNA molecules to DNA sequences.

Over 30 years ago, we found that DNA–silver complexes could be separated by equilibrium centrifugation in Cs_2SO_4 density gradients according to the frequency of silver-binding sites on the DNA molecules, thus allowing high-resolution fractionation (Corneo et al., 1968). In turn, DNA fractionation using sequence-specific ligands, such as silver ions [or BAMD, 3,6-bis(acetato mercurimethyl)-1,4-dioxane, in later studies], led to the discovery of the striking and unexpected *compositional heterogeneity* of high molecular weight, "main band" (non-satellite, non-ribosomal) bovine DNA (Filipski et al., 1973). Subsequently, it also led (Thiery et al., 1976; Macaya et al., 1976) to three main conclusions: (i) vertebrate genomes are

* Tel.: +39-081-5833215; fax: +39-081-2455807.
E-mail address: bernardi@alpha.szn.it (G. Bernardi)

0378-1119/00/$ - see front matter
PII: S0378-1119(99)00485-0

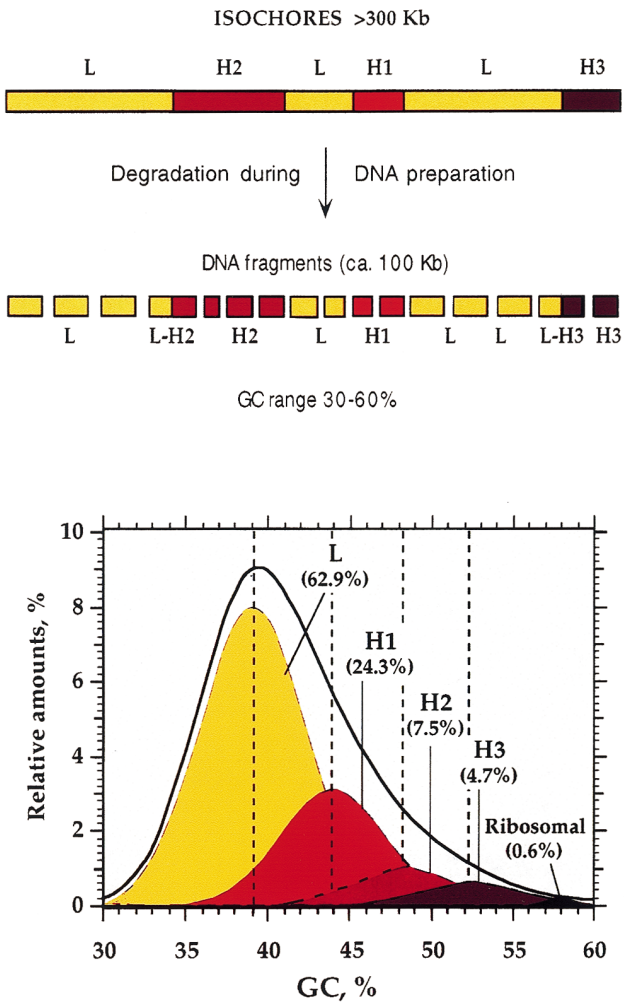

Fig. 1. (Top) Scheme of the isochore organization of the human genome. This genome, which is typical of the genome of most mammals, is a mosaic of large (≫300 kb, on average) DNA segments, the isochores, which are compositionally homogeneous (above a size of 3 kb) and can be partitioned into a number of families. Isochores are degraded during routine DNA preparations to fragments of approx. 100 kb in size. The GC-range of the isochores from the human genome is 30–60% (from Bernardi, 1995). (Bottom) The CsCl profile of human DNA is resolved into its major DNA components, namely the families of DNA fragments derived from isochore families L (i.e., L1+L2), H1, H2, H3. Modal GC levels of isochore families are indicated on the abscissa (broken vertical lines). The relative amounts of major DNA components are indicated. Satellite DNAs are not represented (from Zoubak et al., 1996).

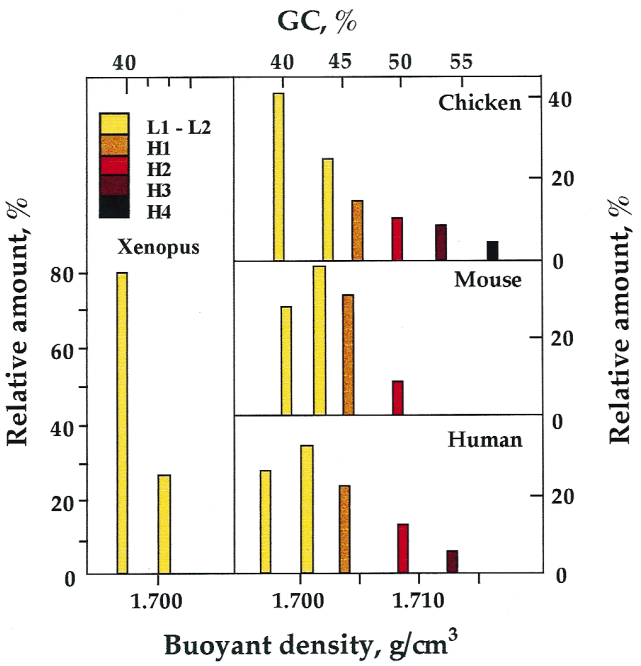

Fig. 2. Compositional patterns of vertebrate genomes. Histograms showing the relative amounts, modal buoyant densities and modal GC levels of the major DNA components (the families of DNA fragments derived from different isochore families; see Fig. 1); from *Xenopus*, chicken, mouse and man, as estimated after fractionation of DNA by preparative density gradient in the presence of a sequence-specific DNA ligand (Ag^+ or BAMD). Satellite and minor DNA components (such as ribosomal DNA) are not shown. (Modified from Bernardi, 1995.)

mosaics of *isochores* (Fig. 1), namely of long DNA segments (≫300 kb), that are compositionally homogeneous (above a certain size originally assessed as 3 kb) and belong to a small number of families characterized by different GC levels (GC is the molar ratio of guanine+cytosine in DNA); (ii) the relative amounts of DNA in the isochore families define the *isochore pattern* of a genome (see below); this *compositional pattern* can be investigated using the large DNA fragments (approx. 100 kb in size) that result from routine DNA preparations; and (iii) isochore patterns are remarkably different in cold- and warm-blooded vertebrates.

In the case of the human genome, the isochore pattern is characterized (Fig. 1) by GC-poor, 'light', L, isochores that represent about 63% of the genome, whereas GC-rich, 'heavy', H1, H2 and H3 isochores make up about 24%, 7.5% and 4.7%, respectively, of the genome, the remaining DNA corresponding to satellite and ribosomal sequences (Bernardi et al., 1985; Zerial et al., 1986; Zoubak et al., 1996).

Another type of compositional pattern is that of coding sequences. In this case, either their GC levels or, more informatively, GC_3, the GC level of their third codon positions, define the pattern (see, for example, the human GC_3 pattern in Fig. 4). Interestingly, the GC_3 patterns of *Xenopus*, chicken, mouse and human mimic the isochore patterns of Fig. 2 (see Bernardi, 1995).

Compositional patterns have been called *genome phenotypes* (Bernardi and Bernardi, 1986), because they differ not only among vertebrate classes, but may also differ among orders within a class, and even among different families within an order (Bernardi et al., 1985; see also Bernardi, 1995, for a review).

It should be noted that the analysis of long sequences from data banks (Ikemura and Aota, 1988; Ikemura

et al., 1990) and compositional mapping (Bernardi, 1989) of chromosomal bands (Gardiner et al., 1990; Krane et al., 1991; Bettecken et al., 1992; Pilia et al., 1993; De Sario et al., 1996, 1997) have provided examples of human isochores and showed sizes ranging from 0.2 Mb (megabases) to several Mb for the longest ones (De Sario et al., 1996, 1997). Sharp isochore boundaries have been demonstrated (Fukagawa et al., 1995; Stephens et al., 1999). These were expected on the basis of the fact that DNA molecules from different isochore families could be separated from each other. Smooth transitions among isochores would have made such separation almost impossible.

3. Compositional correlations

An obvious question is whether there is any correlation between the compositional patterns of coding sequences (which represent as little as 3% of the genome in vertebrates) and the compositional patterns of DNA fragments (97% of which are formed by intergenic sequences and introns). Another question is whether there is any correlation within genes between the composition of the exons and that of the introns. The answer to both questions is yes.

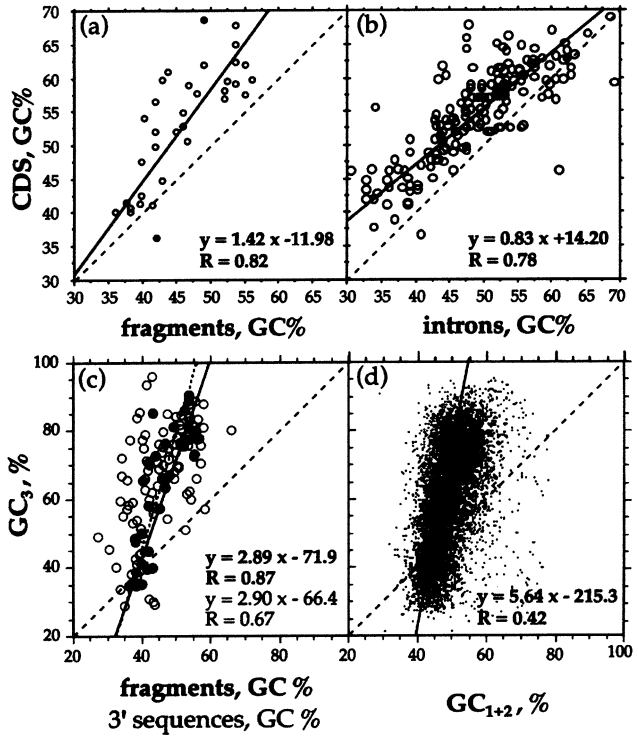

Fig. 3. Correlation between GC levels of human coding sequences and (a) the GC levels of the large DNA fragments in which sequences were localized, or (b) the GC levels of the corresponding introns (top frames). The bottom frames show the correlations between GC_3 of human coding sequences and (c) the GC levels of the DNA fractions in which the genes were localized (filled circles) and of 3′ flanking sequences further than 500 bp from the stop codon (open circles; the solid and the broken lines are the regression lines through the two sets of points); or (d) GC_1+GC_2 values of human sequences. Diagonals (unity slope lines) are also shown (from Clay et al., 1996).

Indeed, linear correlations hold between the GC levels (and the GC_3 levels) of coding sequences and the GC levels of the isochores in which coding sequences are located (see Fig. 3a, c). Interestingly, GC-poor coding sequences and their flanking sequences show very similar values, whereas GC-rich coding sequences are increasingly higher above the diagonal, essentially because GC_3 values depart more and more from the intergenic sequences (Fig. 3c).

Linear correlations (Fig. 3b) also hold between the GC levels of coding sequences and the GC levels of the introns of the same genes (Bernardi et al., 1985; Aïssani et al., 1991; Clay et al., 1996), the GC levels of the former being slightly higher than those of the latter. These differences are much larger in plants (Carels and Bernardi, 2000).

The *genome equations* defining these *compositional correlations*, together with the equations of the *universal correlations* (Fig. 3d) to be further discussed later, amount to a *genomic code* (Bernardi, 1990, 1993a).

As a final remark, one should note that the correlations of Fig. 3a and b are practically the same in the chicken genome (Musto et al., 1999), and possibly in other vertebrate classes.

4. Gene distribution and gene spaces

The correlation between GC_3 levels of coding sequences and GC levels of isochores (Fig. 3c) is especially important, because it allows the positioning of the distribution profile of coding sequences relative to that of DNA fragments, the CsCl profile. In turn, this allowed us to estimate the relative gene density by dividing the percentage of genes located in given GC intervals by the percentage of DNA located in the same intervals. Since it had been tacitly assumed that genes were uniformly distributed in eukaryotic genomes, it came as a big surprise that the *gene distribution* in the human genome (and, for that matter, in the genomes of all vertebrates; see below) is strikingly non-uniform (Fig. 4), gene concentration increasing from a very low average level in L isochores to a 20-fold higher level in H3 isochores (Bernardi et al., 1985; Mouchiroud et al., 1991; Zoubak et al., 1996).

The existence of a break in the slope of gene concentration at 60% GC_3 of coding sequences and at 46% GC of isochores (see Fig. 4) defines two '*gene spaces*' in the human genome. In the '*genome core*' (Bernardi, 1993a, 1995), formed by isochore families H2 and H3 (which make up 12% of the genome), gene concentration is very high (one gene per 5–15 kb) and comparable to

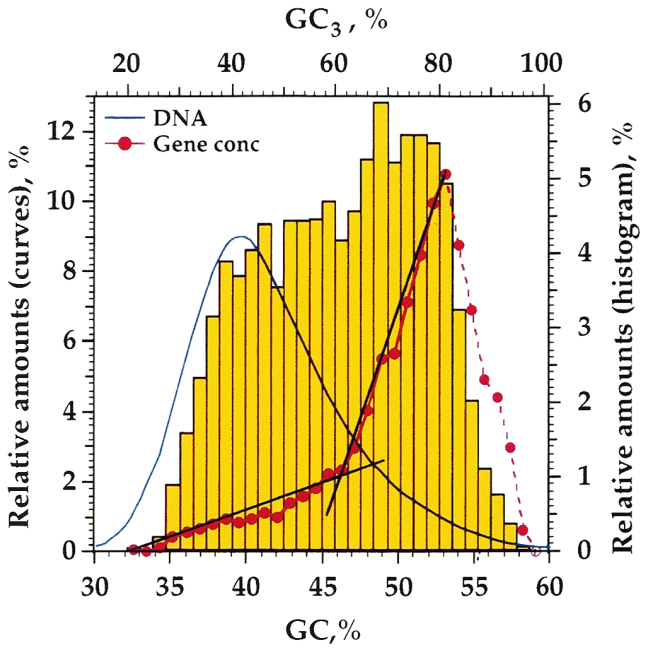

Fig. 4. Profile of gene concentration (red dots) in the human genome, as obtained by dividing the relative numbers of genes in each 2% GC_3 interval of the histogram of gene distribution (yellow bars) by the corresponding relative amounts of DNA deduced from the CsCl profile (blue line). The positioning of the GC_3 histogram relative to the CsCl profile is based on the correlation of Fig. 3c. The apparent decrease in the concentration of protein-encoding genes for very high GC values (broken line) is due to the presence of ribosomal DNA in that region. The last concentration values are uncertain because they correspond to very low amounts of DNA (from Zoubak et al., 1996).

those of compact genomes of higher eukaryotes, whereas in the '*empty space*', formed by isochore families L and H1 (which make up 88% of the genome), gene concentration is very low (one gene per 50–150 kb). Note that the definition of 'genome core' is prompted not only by the position of the break in the gene concentration curve of Fig. 4, but also by the proximity of gene concentrations in L/H1 and H2/H3 isochores, respectively (Fig. 5), by the identical chromosomal distribution of H2 and H3 isochores (Saccone et al., 1996, 1999) and by the similarity of the heptanucleotide comprising the AUG initiation codon of human genes located in L/H1 and H2/H3 isochores, respectively (Pesole et al., 2000).

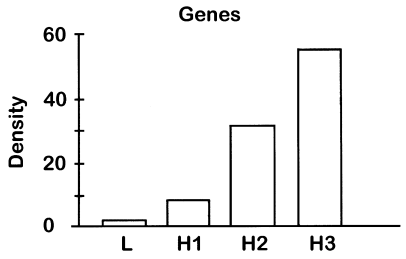

Fig. 5. Density of gene sequences in isochore families. Relative numbers of sequences over relative amounts of isochore families are presented in the histograms (from Zoubak et al., 1996).

About 54% of human genes are located in the small genome core, the remaining 46% being located in the large empty space.

The two gene spaces are characterized by a number of different structural and functional properties. Indeed, most genes located in the genome core (which comprise the majority of housekeeping genes; Larsen et al., 1992) are associated with CpG islands (Aïssani and Bernardi, 1991a,b; Jabbari and Bernardi, 1998), are actively transcribed, and correspond to an *open chromatin structure* (Kerem et al., 1984). This is characterized by the scarcity, or absence, of histone H1, acetylation of histones H3 and H4, and a larger nucleosome spacing (Tazi and Bird, 1990). H2 and H3 isochores colocalize on human metaphase chromosomes (at a 400 band resolution) in two small sets of R (Reverse) bands, the $H3^+$ and $H3^*$ bands (Saccone et al., 1992, 1993, 1996). $H3^+$ bands, which show the strongest hybridization with H3 isochores, undergo very active recombination, and replicate earliest in the cell cycle (Federico et al., 1998). In contrast, GC-poor isochores are located in G (Giemsa) bands and in $H3^-$ R bands (in which H3 isochores cannot be detected), which generally exhibit a closed chromatin structure. Indeed, G and $H3^-$ R bands are characterized by only a few open chromatin regions corresponding to genes that are largely transcribed in a tissue-specific, or in a developmentally regulated manner (Bickmore and Craig, 1997). For this reason, at any given time in development and in any given tissue, the ratio of open to closed chromatin regions is likely to be much higher than the ratio of genes located in the genome core and in the empty quarter. This is an important conclusion as far as the integration of retroviral DNA into the vertebrate genome is concerned, because integration preferentially occurs into open chromatin regions (see Rynditch et al., 1998, for a review). Interestingly, the localization of H3 isochores has now been precisely defined (Fig. 6) in prometaphase chromosomes at a resolution of 850 bands (Saccone et al., 1999). These results deserve at least three comments: (i) some chromosomes (17, 19, 22) have a very high percentage of $H3^+$ bands, whereas other ones (4, 13, 15, 18, X and Y) have a very small percentage or no $H3^+$ bands at all; no explanation is available, so far, for such an uneven distribution of the gene-richest regions of the chromosomes; (ii) the gene-richest bands tend to be far away from centromeres, with many bands in telomeric locations, coinciding, therefore, with the regions endowed with the highest recombination rates; (iii) H2 and H3 isochores, which represent 12% of the genome, colocalize on $H3^+$ bands, which represent 17% of the total genome; since the coverage of the bands by hybridization signals tends to be overestimated, one should conclude that most of the DNA from $H3^+$ bands at the 850-band resolution belongs to isochores H2 and H3.

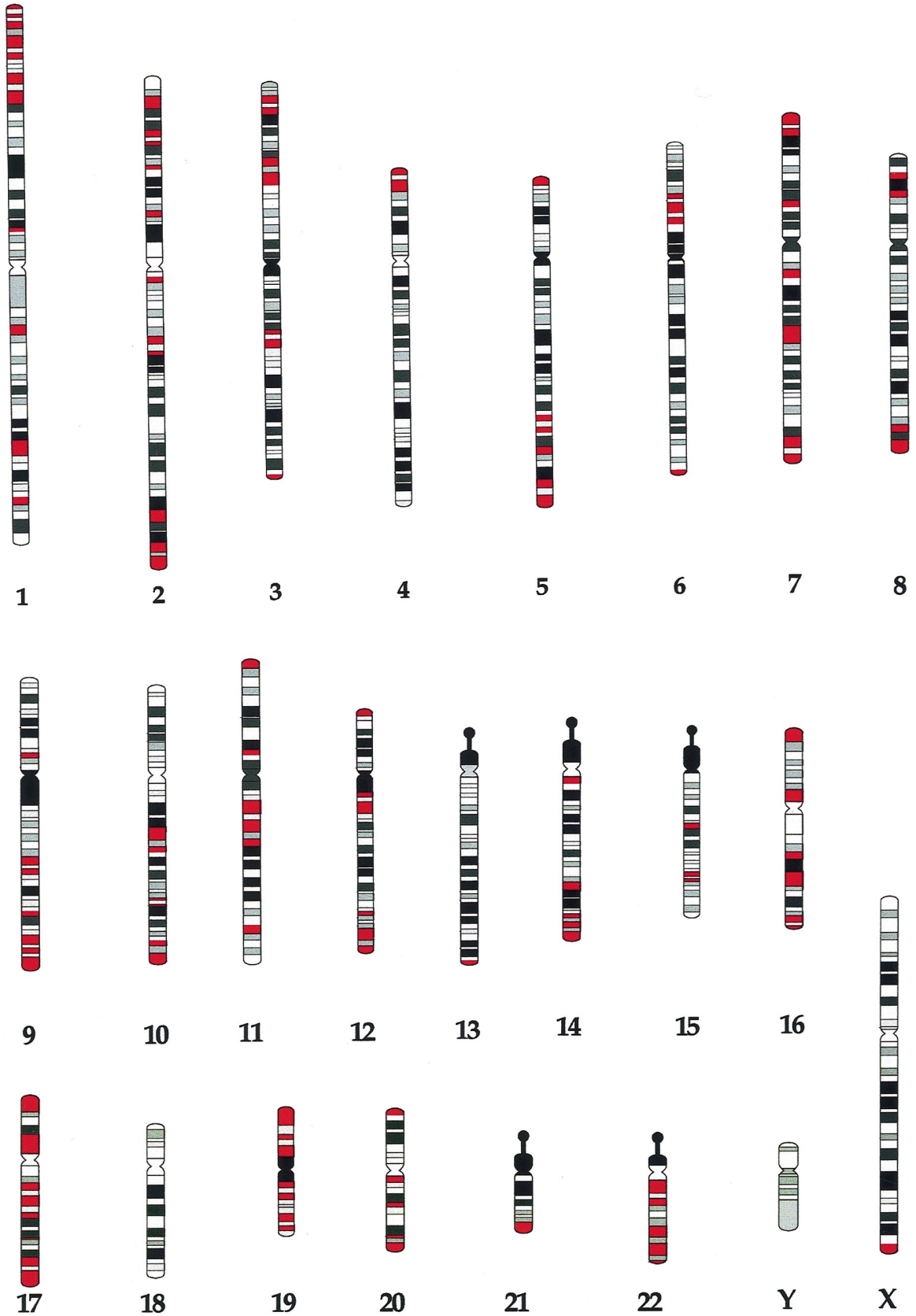

Fig. 6. Ideogram of human chromosomes at a 850 band resolution (Francke, 1994) showing the $H3^{+}$ bands as red bands (from Saccone et al., 1999). Black and grey bands are G(Giemsa) bands, white and red bands are R(Reverse) bands.

Last but not least, the differences in GC_3 among genes located in the genome core and in the empty space lead to a strong bias in codon usage; indeed, at the limit value of 100% GC_3, which is approached by a number of human genes (see Fig. 4), only 50% of the codons are available. Moreover, they also lead to differences in GC_2 and GC_1, because of the universal correlations linking these values (see Fig. 3d). As a result, proteins encoded by genes located in the two gene spaces are different in amino-acid composition (D'Onofrio et al., 1991).

The findings just outlined are of interest in that the compositional patterns (the genome phenotype), the genome equations (the genomic code), and the gene distribution define a eukaryotic genome in terms of its structural and functional properties. This replaces the original, purely operational definition of the genome as the haploid chromosome set (Winkler, 1920), which still is the only one presented, in an explicit or implicit form, in current textbooks (see Lewin, 1997, for example).

For this reason, these results also represent a breakthrough in the long-standing problem of genome organization of vertebrates. Previous attempts based on DNA re-association studies (Britten and Kohne, 1968), as analyzed by separating single- and double-stranded DNA on hydroxyapatite columns (Bernardi, 1965), could not go beyond the important, yet limited, finding of the existence of repeated sequences in eukaryotic genomes.

Finally, these results raise a question: what is the evolutionary background of the compositional properties and of the gene distribution of the human genome?

5. The major compositional transitions of the vertebrate genomes

The compositional pattern just described for the human genome is basically shared by all warm-blooded vertebrates (Sabeur et al., 1993; Mouchiroud and Bernardi, 1993; Kadi et al., 1993). In contrast, cold-blooded vertebrates are endowed with genomes characterized by a much lower level of compositional heterogeneity and by the fact that, as a general rule, they do not reach the high GC levels attained by the genomes of warm-blooded vertebrates (Thiery et al., 1976; Bernardi and Bernardi, 1990a,b). Genes are, however, not uniformly distributed in these genomes either. Indeed, only the GC-richest 10–15% of the genomes of cold-blooded vertebrates hybridize single-copy DNA from human H3 isochores (Bernardi, 1995; Perani, 1996 and unpublished results). This indicates that in cold-blooded vertebrates as well, there is a genome core located in the GC-richest fractions of the genome, even if these GC-richest fractions are much less GC-rich than the corresponding fractions of the genomes of warm-blooded vertebrates.

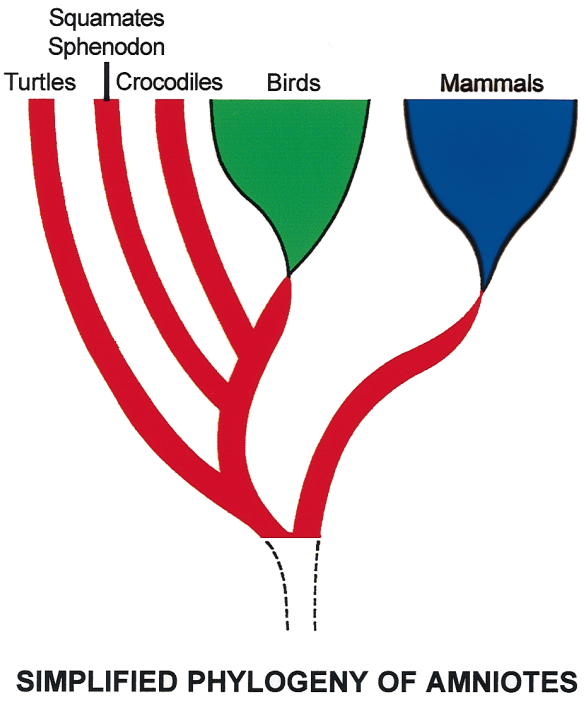

Fig. 7. A simplified phylogeny of Amniotes. The class *Reptilia*, as customarily defined, includes all the red lineages (from Carroll, 1987).

Since mammals and birds originated independently from reptiles (Fig. 7), one should conclude that two major independent compositional genome transitions took place between cold-blooded vertebrates (reptiles) and warm-blooded vertebrates (mammals and birds), and that they concerned a small part of the genome, which is, interestingly, the gene-richest part of it. The majority of the genome which did not undergo the compositional transition had been previously called the *paleogenome*, the minority which did, the *neogenome* (Bernardi, 1989).

The best evidence for the compositional transitions is provided by comparisons of GC_3 values of orthologous genes, as shown initially by Perrin and Bernardi (1987) and Bernardi and Bernardi (1991). When such GC_3 plots are made using genes from human and other mammals sharing the '*general mammalian pattern*' (namely the most widespread mammalian pattern, as defined by the CsCl profile; Sabeur et al., 1993), such as calf (Fig. 8), the regression line passes through the origin and is characterized by a slope of unity, the correlation coefficient being very high. In other words, GC_3 values of orthologous genes of man and calf are very close to each other all over the GC_3 spectrum. If one recalls the existence of compositional correlations between GC_3 and the isochores in which the corresponding genes are located, the human/calf plot indicates that there is a very high degree of similarity of isochore

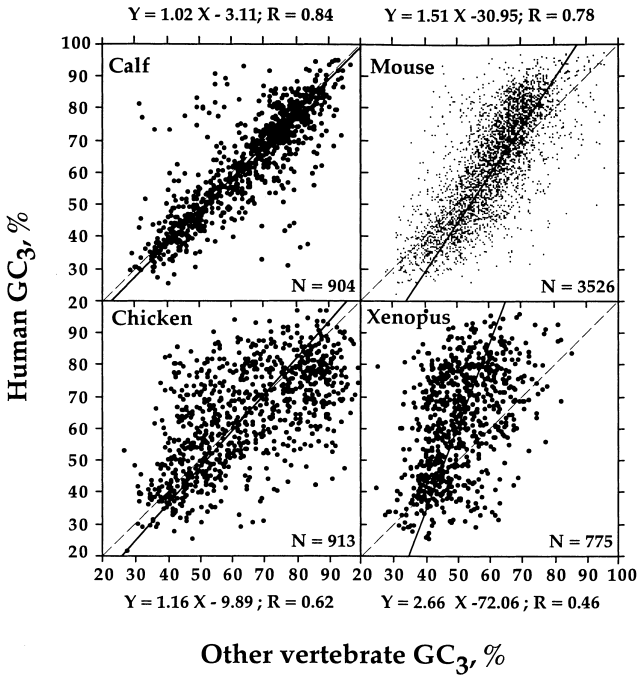

Fig. 8. Correlation between GC_3 values of orthologous genes from human and calf, human and mouse, human and chicken, and human and *Xenopus*. The orthogonal regression lines are shown together with their equations and the diagonal (broken) lines. N is the number of gene pairs.

patterns in the genomes of mammals sharing the general pattern, in agreement with the DNA fractionation results (Thiery et al., 1976; Sabeur et al., 1993). It should be mentioned, however, that some '*special mammalian patterns*' also exist (Salinas et al., 1986; Sabeur et al., 1993; Mouchiroud and Bernardi, 1993; Robinson et al., 1997). So far, however, the only well-demonstrated special pattern is the murid pattern. In this case, if a plot is made using GC_3 values from human and mouse, the correlation coefficient still is very high, but the slope is higher than unity, since GC_3 levels of human genes are higher than the orthologous genes of mouse for GC-rich genes, whereas they are lower for GC-poor genes (Fig. 8).

When the GC_3 plot compares values of orthologous genes from human and chicken (Fig. 8), the slope is close to unity, and the straight line practically coincides with the diagonal, but the correlation coefficient is lower (due to the larger scatter of points, especially in the GC-rich range), yet still highly significant. The human/chicken plot stresses three important points: (i) largely the same genes were affected by the compositional transition in the two independent lines of mammals and birds; (ii) these genes largely have a similar distribution in the isochores of mammals and birds; and (iii) this similar distribution reflects that of their common reptilian ancestor, so providing another argument in favor of the similarity of gene distributions in vertebrates.

In contrast, when the human/*Xenopus* plot was investigated (Fig. 8), points were scattered around the diagonal in the low-GC range, showing no directional change between the two species, whereas human gene values were increasingly higher, on the average, than the corresponding *Xenopus* gene values, as increasingly higher GC_3 values were explored. This results in a slope, 2.7, which is much higher than unity. Interestingly, the correlation coefficient still is very significant.

At this point, one should note that the compositional transitions just described (i) essentially concerned the GC-rich genes from the gene-dense regions, namely from the genome core; (ii) occurred (and were similar) in the independent ancestral lines of mammals and birds, but in no cold-blooded vertebrate (the genomes of very few cold-blooded vertebrates have relatively GC-rich genomes, but these are compositionally homogeneous; see Bernardi and Bernardi, 1990a,b); and (iii) stopped at least with the appearance of present-day mammals and birds, as indicated by the essentially identical patterns found in different mammalian orders (like primates and artiodactyls; see Fig. 8), as well as in different avian orders (Kadi et al., 1993; Bernardi et al., 1997); it should be noted that mammalian orders diverged from a common ancestor according to a star-like phylogeny (see Fig. 12, bottom right), some 100 Myr ago. This means that the *convergent compositional evolution* undergone by the genome core of the ancestors of mammals and birds reached a *compositional equilibrium*, at least as early as the time 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. This conservation is remarkable, if one considers that about 50% of the human genes underwent a compositional transition.

Two additional points should be made here. The first one is that the compositional transitions involved more than the compositional changes just described. Indeed, (i) DNA methylation and CpG doublet concentration decreased by a factor of two between fishes/amphibians and mammals/birds (Jabbari et al., 1997); (ii) CpG islands, namely regulatory sequences about 1 kb in size, located 5′ of GC-rich genes, rich in GC and in unmethylated CpG doublets, were formed (Bernardi, 1989; Aïssani and Bernardi, 1991a,b); and (iii) T bands appeared in metaphase chromosomes; at the same time karyotype changes and speciation increased (Bernardi, 1993b).

The second point is that the genome organization of reptiles deserves additional work. While the reptilian genomes studied so far show the compositional pattern of cold-blooded vertebrates and do not show CpG islands (Thiery et al., 1976; Bernardi and Bernardi, 1990a,b; Aïssani and Bernardi, 1991a,b), the genome size and the methylation levels are more similar to those of DNAs from warm- than to those of DNAs from

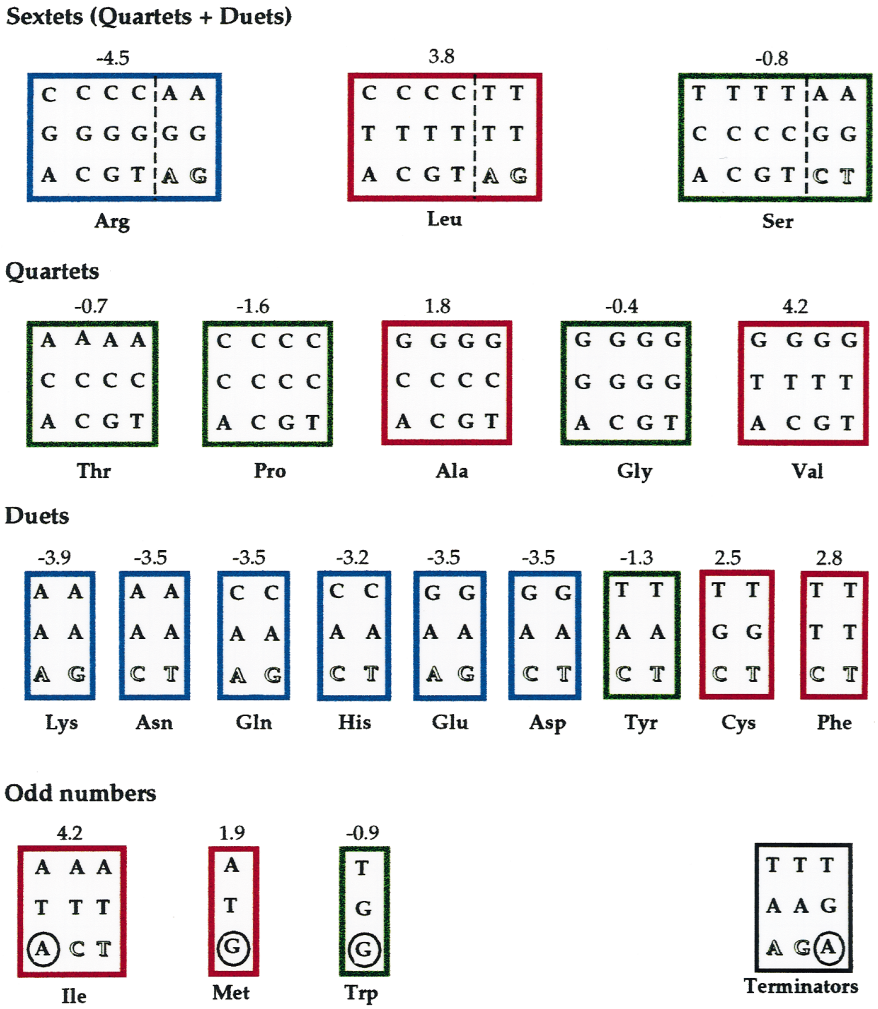

Fig. 9. The Grantham (1980) representation of the genetic code was modified in that codons rather than anticodons are shown, a distinction is made among third position nucleotides of quartet, duet and odd number codons, and hydropathy values for amino acids using the scale of Kyte and Doolittle (1982) are shown. Red, blue and green boxes indicate the most hydrophobic, the most hydrophilic and the intermediate amino acids, respectively (from D'Onofrio et al., 1999a,b).

cold-blooded genomes (Jabbari et al., 1997). Very little information is available about reptilian coding sequences. The very recent claim that reptiles exhibit a warm-blooded isochore structure (Hughes et al., 1999) hinged on the demonstration of a unity slope in the GC_3 plot of reptiles vs. birds. Unfortunately, the very small sample of partial coding sequences from only two reptiles that was used (ten sequences from a crocodile and six from a turtle), cannot define a statistically reliable slope. Moreover, there is no evidence that the isochore patterns of crocodiles and turtles are the same as those of the reptilian ancestors of mammals and birds.

6. The causes of compositional transitions in vertebrate genomes

An obvious question concerns the cause(s) (i) of the compositional genome transitions; and (ii) of the maintenance of the new compositional patterns. The original explanation for the compositional transition (Bernardi and Bernardi, 1986) was that *natural selection* was responsible. Natural selection, the differential multiplication of mutant types, occurs through the elimination of organisms with deleterious mutations (*negative selection*) and, very rarely, via the preferential propagation of organisms with advantageous mutations (*positive selection*).

Compositional changes in genomes obviously are the result of the interplay of many factors which cannot be easily identified. We thought, however, that, if present, a major factor could be recognized in the case of changes affecting a small homogeneous taxon, such as vertebrates. As a working hypothesis, we therefore proposed that the selective advantages provided by the compositional genome transitions to the vertebrates that are characterized by high body temperatures were the higher *thermodynamic stabilities* of DNA, of RNA, and of the

important proteins (e.g., housekeeping proteins) encoded by the GC-rich coding sequences, all these advantages being achieved simultaneously (see, however, Section 10).

As far as DNA stability is concerned, it should be recalled that H3 and H2 isochores have their highest concentration in a set of R bands (Saccone et al., 1992, 1993, 1996) that largely coincide with chromosomal bands previously identified as particularly resistant to thermal denaturation (Dutrillaux, 1973). As for RNA stability, 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).

Protein stability requires a more detailed discussion. We should recall that universal correlations were reported (D'Onofrio et al., 1991; D'Onofrio and Bernardi, 1992) between the GC levels of the three codon positions. These correlations hold both intra-genomically (among the coding sequences of the human genome; see Fig. 3d, for example) and inter-genomically (among the coding sequences of different genomes, as averaged per genome, or per isochore family, in the case of compartmentalized genomes). These correlations have recently been re-analyzed and confirmed using a much larger data set (D'Onofrio et al., 1999b). Two implications of the positive correlations between GC_3 and GC_1, or between GC_3 and GC_2, are that GC_3 increases should be accompanied (i) by increases of amino acids encoded by codons of the GC class (the codons having G and/or C in first and second positions; D'Onofrio et al., 1991) and by decreases of amino acids encoded by codons of the AT class (the codons having A and/or T in first and second positions); and (ii) by increases in quartet codons and decreases in duet codons, simply because the former are GC-richer than the latter in first and second positions. Indeed, the first point was verified (D'Onofrio et al., 1991), and rechecked on very recent data (D'Onofrio et al., 1999a, b); and the second point was also found to be true (not shown; see, however, Fig. 9, which displays the Grantham representation of the genetic code).

A more recent finding (D'Onofrio et al., 1999a,b) is that in both bacteria and vertebrates GC_3 is positively correlated not only with the frequency of amino acids of the GC class, but also with the hydrophobicity of the amino acids (Fig. 10), in agreement with Naylor et al. (1995), but in disagreement with Gu et al. (1998). Interestingly, although the slopes of the two regression lines of Fig. 10 are identical, prokaryotic values are systematically higher than vertebrate values. This difference was accompanied by another remarkable feature of prokaryotic versus eukaryotic proteins, namely by a cysteine level in prokaryotes half as high as that of vertebrates. This suggests that the higher hydrophobicities of prokaryotic proteins were replaced, as a stabilizing

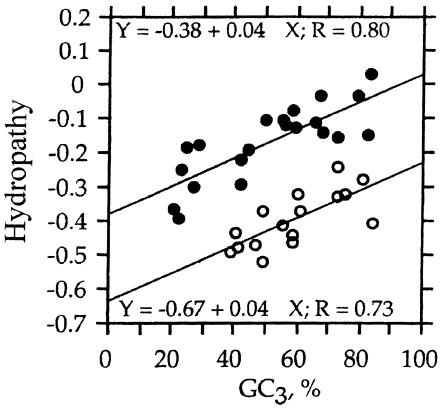

Fig. 10. Correlation between hydropathy values (using the Kyte and Doolittle scale) and GC_3 for amino acids pooled from individual prokaryotic (closed circles) and vertebrate genomes (open dots) (from D'Onofrio et al., 1999a,b).

factor, by the higher levels of disulphide bridges found in eukaryotic proteins. Incidentally, the difference in hydrophobicity of prokaryotic and eukaryotic proteins has the same magnitude as that between proteins encoded by GC-rich and GC-poor proteins in the human genome.

At this point, we analyzed the hydropathies of orthologous proteins from *Xenopus* and man (Cruveiller et al., 1999). Genes were divided into three groups, according to GC_3 levels of human coding sequences. When the hydropathies of amino acids encoded by the three groups were compared (Fig. 11), differences were found to be negligible in the GC_3-poor group and in the intermediate group (as small as those found between homologous proteins of calf and man), whereas they were very pronounced for about half of the amino acids of the GC_3-rich group. Obviously, the increased hydrophobicities of human proteins relative to those of their orthologs from *Xenopus* provided an evidence for the increased stability of the former. These results are

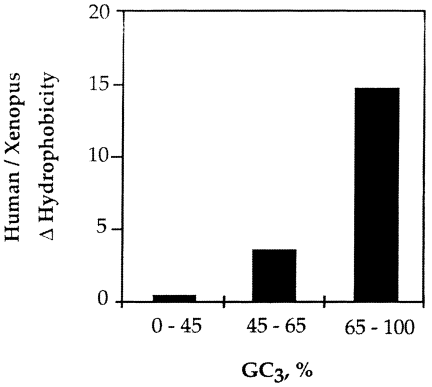

Fig. 11. Difference histogram of hydropathy values of homologous proteins from human and *Xenopus*. Proteins are partitioned in three groups according to the GC_3 values of the corresponding coding sequences (data from Cruveiller et al., 1999).

remarkable if one considers that the total number of amino acid substitutions was the same, about 25%, in the three classes of proteins, and that the directional changes found in the GC_3-rich group concerned as much as 25% of all changes.

7. The maintenance of the compositional patterns of warm-blooded vertebrates

The maintenance of the compositional pattern of mammalian genes was initially investigated by an *intergenic analysis* comparing the average composition of synonymous and non-synonymous positions of orthologous genes. It was found that the frequencies of synonymous substitutions were correlated with the frequencies of non-synonymous substitutions (a point already reported by other authors) and gene-specific (Mouchiroud et al., 1995), suggesting that synonymous and non-synonymous rates are under some common selective constraints. The constraints on synonymous positions are likely to be related to selection for translational accuracy, as suggested by Akashi (1994).

A crucial advance was then made by subjecting the nucleotide substitutions to an *intragenic analysis*. In a first approach, synonymous positions of quartet codons of orthologous coding sequences from four orders of mammals were divided into conserved (no change), intermediate (one change) and variable (more than one change) positions. The three classes of positions in GC-rich genes were shown to deviate significantly from expectations based on a stochastic process in which nucleotide substitutions accumulate at random over time, whereas this was not the case for GC-poor genes (Cacciò et al., 1995). Moreover, synonymous positions (especially conserved positions) of quartet codons from GC-rich coding sequences exhibited significantly different base compositions compared to expectations based on a 'random' substitution process from the 'ancestral' sequence to the present-day sequences; significant differences were, by contrast, rare in GC-poor genes (Zoubak et al., 1995). The latter results not only supported the idea of negative selection for a large number of synonymous positions of GC-rich genes, but also provided a demonstration for the 'neutrality' of changes in most synonymous positions of GC-poor genes and in a number of synonymous positions of GC-rich genes.

A second approach used a window analysis on the same set of orthologous genes (Alvarez-Valin et al., 1998). This showed that the intragenic variability of synonymous rates was correlated with that of non-synonymous rates, and that the variation in GC level (and especially in C level) of all synonymous positions along each gene was correlated with the variation in synonymous rate. These results indicate that not only synonymous and non-synonymous rates, but also GC levels of synonymous positions of GC-rich coding sequences, are under some common selective constraints.

The third approach was to analyze regions of the same set of genes that corresponded to different predicted protein structures. Regions corresponding to α helix, β sheet and coil were shown to be characterized by different synonymous substitution rates, and by different levels of GC and of individual nucleotides (Chiusano et al., 1999; see Table 1). More recently, the same analysis was performed (Alvarez-Valin et al., 2000) on 19 *Leishmania* genes encoding the surface metalloproteinase GP63, whose crystallographic structure is known. A sliding window analysis showed (i) that the rate of synonymous substitutions along the GP63 gene is highly correlated with both the rate of amino acid substitution and codon usage; and (ii) that there is a clear relationship between rates and the tertiary structure of the encoded protein, 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.

The results just mentioned convincingly demonstrate that negative selection was responsible for the maintenance of the new compositional patterns. Indeed, it is impossible to account for those findings using alternative explanations that have been proposed (see following section).

It should be noted that negative selection could also explain the compositional transitions, if one assumes that the threshold of optimal GC values increased with increasing body temperature from cold- to warm-blooded vertebrates (Bernardi, 1993c).

Compositional changes and their maintenance during vertebrate evolution may be summarized as indicated in Fig. 12. GC_3 values of reptilian coding sequences are subdivided into a GC-poor class (indicated as $GC_3 \sim 50\%$) and a GC-rich class (indicated as $GC_3 > 50\%$). The GC-poor class (comprising the genes of the paleogenome) did not undergo directional changes. The simplest explanation is that, in this case,

Table 1
Base composition, synonymous and non-synonymous substitution rates in coding sequence regions corresponding to different secondary structures of proteins[a]

	Coil	α-Helix	β-Sheet
A_3	0.18	0.15	0.09
T_3	0.19	0.15	0.19
C_3	0.38	0.32	0.45
G_3	0.27	0.40	0.28
GC_3	0.65	0.72	0.73
SYN	0.56	0.51	0.41
NSYN	0.13	0.12	0.09

[a] The values presented are mean values; see Chiusano et al. (1999) for standard deviations and *P* values.

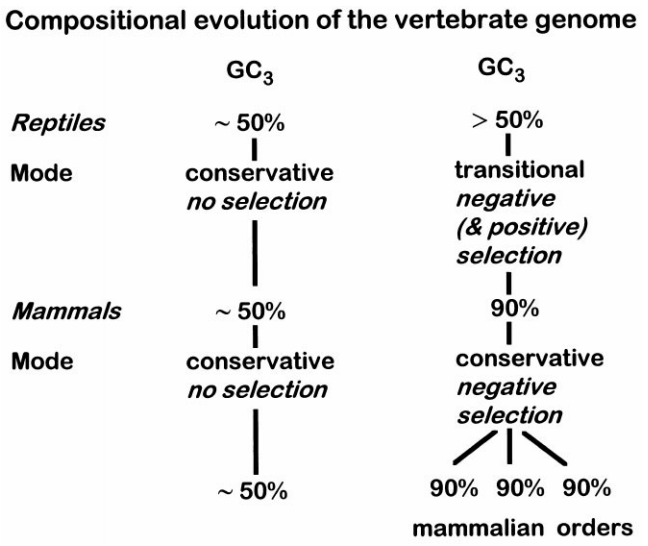

Fig. 12. Scheme of the compositional evolution of the amniote genome, reptiles to mammals (see text).

evolution proceeded according to a conservative mode, simply because base substitutions were stochastic (Zoubak et al., 1995).

The GC-rich class underwent a compositional transition to high GC_3 values (arbitrarily indicated as 90% GC_3 in Fig. 12) by negative (and, to a very small extent, positive) selection. The newly attained values were maintained by negative selection as suggested by very close GC_3 values of orthologous genes from mammalian orders belonging to the general mammalian pattern.

8. Alternative explanations: mutational bias

Several alternative explanations have been proposed to account for the compositional transitions and their maintenance. These comprise biases in DNA repair (Filipski, 1987), mutational bias (Sueoka, 1988), changes in nucleotide pools during DNA replication (Wolfe et al., 1989) and recombination (Eyre-Walker, 1993). Since biases in DNA repair, changes in nucleotide pools and recombination have already been ruled out as valid explanations (see Bernardi et al., 1988, 1993; Mouchiroud et al., 1995), the discussion here will be limited to mutational bias which, at present, seems to be the most favored alternative interpretation.

A mutational bias was originally proposed to be responsible for the broad compositional spectrum of bacterial genomes (Freese, 1962; Sueoka, 1962). The existence of mutational biases cannot be denied, since they are abundantly demonstrated by mutator mutations in bacteria. What can be questioned, however, is that they are the cause of the compositional changes in the genomes of vertebrates and of the conservation of the changes. Indeed, one should not forget that mutator mutations are due to changes in some sub-units of the replication machinery, and that a single replication machinery exists in the nucleus of vertebrates. Under the mutational bias explanation, changes should, therefore, affect the totality of the genome. It is then difficult to understand (i) why the compositional changes leading to the formation of GC-rich isochores occurred in only a very small part of the vertebrate genome, i.e., why the changes were regional; (ii) why these regional changes never were detected in any cold-blooded vertebrate; (iii) why they largely paralleled each other in the independent lines of mammals and birds; (iv) why they reached an equilibrium a long time ago, both in mammals and birds, and were maintained since; and, most importantly, (v) why they were different in exons and introns of the same genes and in their flanking regions (CpG islands, 5′ and 3′ untranslated sequences; Pesole et al., 2000), as well as in different regions of the coding sequences that correspond to specific protein structures (Chiusano et al., 1999; Alvarez-Valin et al., 2000).

It should be noted that an additional independent argument against mutational bias and in favor of selection has been made using population genetics data from the major histocompatibility genes of several mammalian species (Eyre-Walker, 1999). Finally, an analysis of the compositional evolution of α and β globin pseudogenes leading to the peremptory conclusion that isochores derive from mutation not selection (Francino and Ochman, 1999) has been shown to be inconclusive (Eyre-Walker, personal communication).

9. Objections to selection

While a mutational bias can be ruled out as the cause of the formation and maintenance of GC-rich isochores in warm-blooded vertebrates (a point further stressed by results concerning the murid pattern; see below), objections have been raised against the selectionist interpretation. They can, however, be answered.

(i) The low GC levels of some thermophilic bacteria do not contradict, as claimed (Galtier and Lobry, 1997), the selectionist interpretation discussed above. Indeed, different strategies were apparently developed by different organisms to cope with long-term high body temperatures. It is now known that the DNAs of such thermophilic bacteria are very strongly stabilized by particular DNA-binding proteins (Robinson et al., 1998) and that, in turn, their proteins can be stabilized by thermostable chaperonins (Taguchi et al., 1991).

(ii) Another anti-selectionist argument has been seen in the fact that changes in GC-rich isochores concerned not only coding sequences, but also non-coding intragenic (introns) and intergenic sequences. There are, however, several reasons to believe that selection can and does operate on non-coding sequences of the

genome core. Indeed, one should not forget that introns of GC-rich genes are typically very short in both vertebrates (Duret et al., 1995) and plants (Carels and Bernardi, 2000) and may be endowed with regulatory roles (in the mammalian α-globin gene a promoter is located in an intron); and that intergenic sequences of the genome core also are very short (because of the high density of genes), comprise regulatory sequences and CpG islands, and seem to have a role in the expression of flanking genes. Indeed, the transcription of integrated retroviral sequences (which belong into a GC-rich and a GC-poor compositional classes; Zoubak et al., 1992) is optimal when these sequences are located in compositionally matching chromosomal environments of the host, in other words when their location is the same as that of compositionally similar host genes. In contrast, they are not transcribed when they are located in a non-matching environment (see Rynditch et al., 1998, for a review).

(iii) Yet another anti-selectionist argument was seen in the fact that some duplicated genes, like the human GC-poor β globin and GC-rich α globin genes, are expressed in the same cells and fulfill the same role (Wolfe et al., 1989). This argument neglects, however, the fundamentally different regulation of these two genes, the latter (in contrast to the former) having an internal promoter and being associated with a CpG island instead of a TATA box.

(iv) The anti-selectionist argument that the population size of mammalian orders is too small to allow the negative selection process to take place in the time available, since the appearance of mammals is a weak one if one considers that the time taken for the compositional transition to occur may have been very long, anything between 120 Myr (the time difference between the appearance of early mammals, −220 Myr, and that of present-day mammals, −100 Myr) and 220 Myr (the time difference between the appearance of mammalian-like reptiles, −320 Myr and that of present-day mammals; times are from Carroll, 1987).

(v) The anti-selectionist point (iv) is also contradicted by the 'minor shift' between the general mammalian pattern and the murid pattern. This consists in that the GC-richest and GC-poorest genes of the murids are less GC-rich and less GC-poor, respectively, compared with the orthologous genes of other mammals (see Fig. 8). Now it is known that murids (i) have a defective repair system (see Holliday, 1995) and (ii) exhibit, as a consequence, higher rates of synonymous substitutions (by a factor of 5–10) relative to human coding sequences (Wu and Li, 1985; Gu and Li, 1992), and (iii) are endowed with a pattern which is derived from the general mammalian pattern (Galtier and Mouchiroud, 1998). Under the mutational bias hypothesis, this should lead to an increased bias in composition relative to the general mammalian pattern from which the murid pattern is derived. The opposite is found, however, since the higher mutational input clearly randomizes the composition of synonymous positions, as well as of other sequences, leading for instance, to an 'erosion' of CpG islands (Aïssani and Bernardi, 1991a, b; Matsuo et al., 1993). This provides an additional, independent argument against the mutational bias being responsible for the maintenance of isochores.

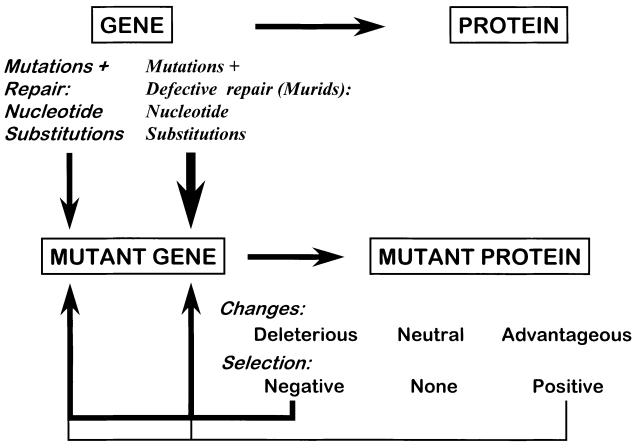

Fig. 13. Scheme of the equilibrium between mutational input (nucleotide substitutions) and negative selection. The equilibrium on the left is that of genes from mammals showing the general mammalian pattern, the equilibrium on the right is the case of genes from murids (see text).

There is, however, a more important implication of the 'minor shift'. As already mentioned, the very similar compositional patterns of genomes from mammalian orders that were separated for 100 million years indicates that this pattern was already present in the common ancestor of present-day mammals. This common pattern clearly was the result of an equilibrium between the (essentially random) mutational input and negative selection. When the mutational input underwent an increase, as in the case of murids, a new equilibrium was reached; see Fig. 13 (as witnessed by the very similar compositional pattern of different murids). This new equilibrium already existed in the common ancestor of murids and can, therefore, be dated at less than 30 million years ago, the time of the divergence of murids, that have a defective repair system, from other rodents that share the general mammalian pattern. This means that mutational input and negative selection could reach a different equilibrium in less than this time interval and rules out the objection that population sizes of mammalian orders were too small to allow negative selection to take place.

10. Conclusions

In conclusion, recent results from our laboratories support the original working hypothesis (Bernardi and

Bernardi, 1986) that natural selection underlies the regional compositional changes accompanying the transition from cold- to warm-blooded vertebrates and maintain the novel, high GC levels attained (Cacciò et al., 1995; Zoubak et al., 1995; Alvarez-Valin et al., 1998, 2000; Chiusano et al., 1999). They considerably refine the original idea and have led to some new insights.

1. There is an equilibrium between mutational input and natural selection. The most straightforward evidence for the equilibrium is provided by the extremely close GC_3 levels of orthologous genes of mammals from orders that have been separated for 100 million years. It is obvious that maintaining the high GC_3 levels of GC-rich genes cannot be accomplished without negative selection of changes towards lower GC_3 levels. The second evidence is that, when the mutational input is considerably increased because of a defective repair, as in the case of murids, the equilibrium is shifted towards less extreme GC_3 values (see Figs. 8 and 13).
2. Synonymous and non-synonymous rates are correlated. This can be understood in the light of the role played by third codon positions in so far as speed and accuracy of translation is concerned. Functionally important amino acids are conserved and so are the third positions of their codons. Under these circumstances, it is not surprising that both synonymous and non-synonymous rates as well as codon usage are correlated with the secondary and tertiary structure of proteins.
3. Negative selection maintains the compositional pattern of coding sequences from the genome core of warm-blooded vertebrates through the elimination from the population of individuals carrying deleterious changes in the corresponding proteins. The question remains as to how the compositional patterns of the introns and of the intergenic sequences of such coding sequences are maintained. As a working hypothesis, one can suggest that these non-coding sequences, which are extremely short in the genome core, are essentially made up of regulatory sequences (promoters, enhancers) and of sequences whose composition and primary structure is influencing the expression of the contiguous coding sequences. The effect of base composition of the chromosomal environment on the transcription of integrated proviral sequences is in favor of such an explanation. How the compositional patterns of these non-coding sequences were formed at the transition from cold- to warm-blooded vertebrates is a matter of speculation. A possibility is that the compositional changes in the proteins interacting with those sequences determined, in turn, the compositional changes in the latter. An alternative explanation, which is not exclusive of the previous one, and is again suggested by retroviral integration results, is that the formation of GC-rich intergenic and intronic sequences is favored by the transcription of GC-rich coding sequences (as well as by the structural changes in primary RNA transcripts). A more general discussion about these issues and their relationship with the neutral and nearly neutral theories will be presented elsewhere.

Acknowledgements

The author thanks the European Union for a Scholarship from the Senior Research Grant Programme in Japan and Professor Takashi Gojobori for his warm hospitality at the Center for Information Biology, National Institute of Genetics, Mishima 411, Japan, as well as Drs. Takashi Gojobori, Toshimichi Ikemura and Tomoko Ohta for discussions. Thanks are also due to all the co-authors of the primary publications, to Fernando Alvarez-Valin, Marcos Antezana, Giacomo Bernardi, Nicolas Carels, Maria Luisa Chiusano, Oliver Clay, Stephane Cruveiller, Giuseppe D'Onofrio, Concetta Federico, Kamel Jabbari, Giorgio Matassi, Hector Musto, Alla Rynditch and Salvatore Saccone for discussions, to Giovanna Di Gennaro and Rosamaria Sole for secretarial help and to Giuseppe Gargiulo for the art work.

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.

Akashi, H., 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translation accuracy. Genetics 136, 927–935.

Alvarez-Valin, F., Jabbari, K., Bernardi, G., 1998. Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlation. J. Mol. Evol. 46, 37–44.

Alvarez-Valin, F., Tort, J.F., Bernardi, G., 2000. Non-random spatial distribution of synonymous substitutions in the leishmanial GP63 gene. Genetics. submitted for publication.

Bernardi, G., 1965. Chromatography of nucleic acids on hydroxyapatite. Nature (Lond.) 206, 779–783.

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

Bernardi, G., 1990. Le génome des vertébrés: organisation, fonction et evolution. Biofutur 94, 43–46.

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

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

Bernardi, G., 1993c. 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., 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., Ehrlich, S.D., Thiery, J.P., 1973. Deoxyribonucleases: specificity and use in nucleotide sequence studies. Nature New Biol. 246, 36–40.
Bernardi, G., Hughes, S., Mouchiroud, D., 1997. The major compositional transitions in the vertebrate genome. J. Mol. Evol. 44, S44–S51.
Bernardi, G., Mouchiroud, D., Gautier, C., 1993. Silent substitutions in mammalian genomes and their evolutionary implications. J. Mol. Evol. 37, 583–589.
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., 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.
Bettecken, T., Aïssani, B., Müller, C.R., Bernardi, G., 1992. Compositional mapping of the human dystrophin gene. Gene 122, 329–335.
Bickmore, W., Craig, J., 1997. Chromosome bands: patterns in the genome. Molecular Biology Intelligence Unit, R.G. Landes Company, Austin, TX.
Britten, R.J., Kohne, D.E., 1968. Repeated sequences in DNA: hundreds of thousands of copies of DNA. Science 161, 529–540.
Cacciò, 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.
Cacciò, S., Zoubak, S., D'Onofrio, G., Bernardi, G., 1995. Nonrandom frequency patterns of synonymous substitutions in homologous mammalian genes. J. Mol. Evol. 40, 280–292.
Carels, N., Bernardi, G., 2000. Two classes of gene in plants. Genetics, in press.
Carroll, R.L., 1987. Vertebrate Paleontology and Evolution. Freeman, New York, NY.
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.
Clay, O., Cacciò, S., Zoubak, S., Mouchiroud, D., Bernardi, G., 1996. Human coding and non-coding DNA: compositional correlations. Mol. Phylogenet. 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.
Cruveiller, S., D'Onofrio, G., Jabbari, K., Bernardi, G., 1999. Different hydrophobicities of orthologous proteins from *Xenopus* and man. Gene 238, 15–21.
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., Roizès, 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., Alvarez-Valian, F., Cruveiller, S., Bernardi, G., 1999a. Evolutionary genomics of vertebrates and its implications. Ann. NY Acad. Sci. 870, 81–94.
D'Onofrio, G., Jabbari, K., Musto, H., Bernardi, G., 1999b. The correlation of protein hydropathy with the composition of coding sequences. Gene 238, 3–14.
D'Onofrio, G., Mouchiroud, D., Aïssani, B., 1991. Correlations between the compositional properties of human genes, codon usage and aminoacid composition of proteins. J. Mol. Evol. 32, 504–510.
Duret, L., Mouchiroud, D., Gouy, M., 1995. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J. Mol. Evol. 40, 308–317.
Dutrillaux, B., 1973. Nouveau système de marquage chromosomique: les bandes T. Chromosoma 41, 395–402.
Eyre-Walker, A., 1993. Recombination and mammalian genome evolution. Proc. R. Soc. Lond. B. Biol. Sci. 252 (1135), 237–243.
Eyre-Walker, A., 1999. Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics 152, 675–683.
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.
Filipski, J., 1987. Correlation between molecular clock ticking, codon usage fidelity of DNA repair, chromosome banding and chromatin compactness in germline cells. FEBS Lett. 217, 184–186.
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.
Francino, M.P., Ochman, H., 1999. Isochores result from mutation not selection. Nature 400, 30–31.
Franke, U., 1994. Digitized and differentially shaded human chromosome ideograms for genomic applications. Cytogenet. Cell Genet. 65, 206–219.
Freese, E., 1962. On the evolution of base composition of DNA. J. Theor. Biol. 3, 82–101.
Fukagawa, T., et al., 1995. Characterization of the boundary region of long-range G+C% mosaic domains in the human MHC locus: pseudoautosomal boundary-like sequence near the boundary. Genomics 25, 184–191.
Gardiner, K., Aïssani, B., Bernardi, G., 1990. A compositional map of human chromosome 21. EMBO J. 9, 1853–1858.
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.
Grantham, R., 1980. Workings of the genetic code. Trends Biochem. Sci. 5, 327–333.
Gu, X., Hewett-Emmett, D., Li, W.H., 1998. Directional mutational pressure affects the amino acid composition and hydrophobicity of proteins in bacteria. Genetica 102/103, 383–391.
Gu, X., Li, W.H., 1992. Higher rates of amino acid substitution in rodents than in humans. Mol. Phylogenet. 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, S., Zelus, D., Mouchiroud, D., 1999. Warm-blooded isochore structure in Nile crocodile and turtle. Mol. Biol. Evol. in press.
Ikemura, T., Aota, S.I., 1988. Global variation in G+C content along vertebrate genome DNA. J. Mol. Biol. 203, 1–13.
Ikemura, T., Wada, K.N., Aota, S.I., 1990. Giant G+C% mosaic structures of the human genome found by arrangement of genebank human DNA sequences according to genetic positions. Genomics 8, 207–216.
Jabbari, K., Bernardi, G., 1998. CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families. Gene 224, 123–128.
Jabbari, K., Cacciò, S., Païs de Barros, J.P., Desgrès, 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.
Kerem, B.S., Goiten, R., Diamond, G., Cedar, H., Marcus, M., 1984. Mapping of DNAase I sensitive regions of mitotic chromosomes. Cell 38, 493–499.
Krane, D.E., Hartl, D.L., Ochman, H., 1991. Rapid determination of nucleotide content and its application to the study of genome structure. Nucleic Acids Res. 19, 5181–5185.
Kyte, J., Doolittle, R.F., 1982. A simple method for displaying hydropathic character of a protein. J. Mol. Biol. 157, 105–132.
Larsen, F., Gundersen, G., Lopez, R., Prydz, H., 1992. CpG islands as gene markers in the human genome. Genomics 13, 1095–1107.
Lewin, B., 1997. Genes V. Oxford University Press, Oxford.
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.
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.
Mouchiroud, D., Bernardi, G., 1993. Compositional properties of coding sequences and mammalian phylogeny. J. Mol. Evol. 37, 109–116.
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.
Mouchiroud, D., Gautier, C., Bernardi, G., 1995. Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of non-synonymous substitutions. J. Mol. Evol. 40, 107–113.
Musto, H., Romero, H., Zavala, A., Bernardi, G., 1999. Compositional correlations in the chicken genome. J. Mol. Evol. 49, 325–329.
Naylor, G.J.P., Collins, T.M., Brown, W.M., 1995. Hydrophobicity and phylogeny. Nature 373, 565–566.
Perani, P., 1996. Étude de la localisation compositionelle des séquences à copie unique de la famille d'isochore H3 humain et de la séquence télomérique (TTAGGG)n chez les vertébrés à sang chaud. Thesis de Doctorat de l'Université de Paris VII — Denis Diderot. Spécialité: Microbiologie.
Perrin, P., Bernardi, G., 1987. Directional fixation of mutations in vertebrate evolution. J. Mol. Evol. 26, 301–310.
Pesole, G., Bernardi, G., Saccone, C., 2000. Isochore specificity of AUG initiator context of human genes. FEBS Lett. submitted for publication.
Pilia, G., Little, R.D., Aïssani, B., Bernardi, G., Schlessinger, D., 1993. Isochores and CpG islands in YAC contigs in human X26.1-qter. Genomics 17, 456–462.
Robinson, H., Gao, Y., Mccray, B.S., Edmondson, S.P., Shriver, J.W., Wang, A.H.J., 1998. The hyperthermophyle chromosomal protein Sac7d sharply kinks DNA. Nature 392, 202–205.
Robinson, M., Gautier, C., Mouchiroud, D., 1997. Evolution of isochores in rodents. Mol. Biol. Evol. 14 (8), 823–828.
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., Cacciò, S., Kusuda, J., Andreozzi, L., Bernardi, G., 1996. Identification of the gene-richest bands in human chromosomes. Gene 174, 85–94.
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, 11929–11933.
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. Chromosome 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.
Stephens, R., Horton, R., Humphray, S., Rowen, L., Trowsdale, J., Beck, S., 1999. Gene organisation, sequence variation and isochore structure at the centromeric boundary of the human MHC. J. Mol. Biol. 291, 789–799.
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.
Taguchi, H., Konishi, J., Ishii, N., Yoshida, M., 1991. A chaperonin from a thermophilie bacterium *Thermus thermophylus*, that controls refolding of several thermophilic enzymes. J. Biol. Chem. 266, 22411–22418.
Tazi, J., Bird, A., 1990. Alternative chromatin structure at CpG islands. Cell 60, 909–920.
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.
Winkler, H., 1920. Verbreitung und Ursache der Parthenogenesis im Pfanzen- und Tierreich. Fischer, Jena.
Wolfe, K., Sharp, P.M., Li, W.H., 1989. Mutation rates differ among regions of the mammalian genome. Nature 337, 441–456.
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.
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., Clay, O., Bernardi, G., 1996. The gene distribution of the human genome. Gene 174, 95–102.
Zoubak, S., D'Onofrio, G., Cacciò, C., Bernardi, G., 1995. Specific compositional patterns of synonymous positions in homologous mammalian genes. J. Mol. Evol. 40, 293–307.
Zoubak, S., Rynditch, A., Bernardi, G., 1992. Compositional bimodality and evolution of retroviral genomes. Gene 119, 207–213.