Gene, 135 (1993) 57–66

GENE 07566

The isochore organization of the human genome and its evolutionary history – a review*,†

(Chromosomal bands; compositional and gene density maps; genomic code)

Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 75005 Paris, France

Received by E. Zuckerandl: 2 August 1993; Revised/Accepted: 26 August/30 August 1993; Received at publishers: 3 September 1993

SUMMARY

This review will first present some properties (including compositional pattern, correlations between isochores and chromosomal bands, and gene distribution) of the human genome, the most extensively studied among vertebrate genomes. It will then explain how these properties came about during the evolution of the vertebrates.

INTRODUCTION

The word genome is over 70 years old. It was coined by Winkler (1920) to denote the sum total of the genes (of a haploid cell) of an organism. Obviously, the non-coding sequences, whose existence was not known at that time, are now included in the definition. One might think that Winkler's definition is still a valid one because: (*i*) none of the current textbooks of molecular biology or genetics goes any further than it; (*ii*) there is a widespread belief that comprehensive rules about the organization of genomes have not yet emerged; and (*iii*) the various genome projects currently under way have not yet led to new general insights in this field.

Correspondence to: Dr. G. Bernardi, Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2, Place Jussieu, 75005 Paris, France. Tel. (33-1) 4329-5824; Fax (33-1) 4427-7977; e-mail: Bernardi@Arthur.Citi2.FR

*Presented at the COGENE Symposium 'From the Double Helix to the Human Genome: 40 Years of Molecular Genetics', UNESCO, Paris, 21–23 April 1993.

†This paper is dedicated to the memory of Professor Charles Sadron (1902–1993).

Abbreviations: aa, amino acid(s); bp, before present; G-bands, Giemsa (chromosomal) bands; GC, guanine + cytosine (mol%); kb, kilobase(s) or 1000 base pairs; Myr, 10^6 years; R-bands, reverse bands; rDNA, DNA encoding ribosomal RNA; T-bands, telomeric bands.

However, while the original definition obviously is still operationally valid, the genome is much more than the sum total of the coding and non-coding sequences of an organism. The genome is a structural, functional and evolutionary system whose nucleotide sequences obey precise rules that amount to a genomic code (see section **b**). This view has emerged as a result of the compositional approach to the study of the genome, which was developed and used in our laboratory over the past 25 years. The tacit assumption of this approach is that compositional properties (base composition, frequency of short sequences) are important features for genome structure and function. This presentation will review our investigations by taking the isochore organization of the human genome as its focal point, and by making use of its evolutionary history to elucidate some of its properties.

THE HUMAN GENOME: ISOCHORE ORGANIZATION

(a) Isochores, compositional patterns and genome phenotypes

The human genome is a mosaic of isochores (see Bernardi et al., 1985; Bernardi, 1989; 1993a, for reviews), very long (>300 kb, on the average) DNA segments (*i*) that are compositionally homogeneous (above a size of

3 kb; Macaya et al., 1976; see also Bettecken et al., 1992) and (*ii*) that belong to a small number of families which cover a very extended (30–60%) GC range. Random physical and enzymatic degradation occurring during DNA extraction breaks down isochores into the large DNA fragments, 50–100 kb in size, which form routine high-molecular-weight DNA preparations (Fig. 1). It should be stressed that isochores correspond to a size range comprised between those of genes and of chromosomal bands. This size range is poorly known, and yet it is very important for understanding genome organization.

The compositional distribution of large DNA fragments from the human genome is characterized by the presence of five families of fragments (the 'major DNA components') derived (*i*) from two GC-poor isochore families L1 and L2, forming together some 60% of the genome; (*ii*) from two GC-rich isochore families, H1 and H2, forming about 20% and 10% of the genome, respectively; and (*iii*) from one very GC-rich isochore family, H3, forming almost 5% of the genome (Fig. 2). The remaining DNA is formed by satellite DNA and minor DNA components (like rDNA), which may also be considered as isochores because of their compositional homogeneity; these components will, however, be neglected here. The major DNA components were discovered twenty years ago in the bovine genome (Filipski et al., 1973) using Cs_2SO_4 preparative density gradient centrifugation in the presence of a sequence-specific DNA ligand, Ag^+. This approach, previously used in order to fractionate mammalian satellite DNAs (Corneo et al., 1968), has a much higher resolving power than CsCl density gradient centrifugation (compare the top and bottom parts of Fig. 3). Isochores were discovered in 1976

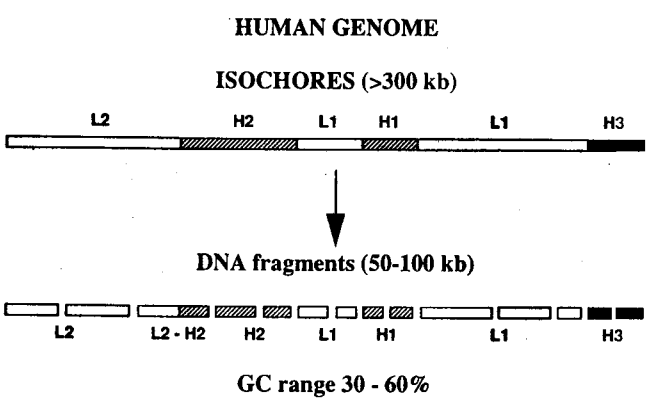

Fig. 1. Scheme of the isochore organization of the human genome. This genome is a mosaic of large (> 300 kb) DNA segments, the isochores, which are compositionally homogeneous (above a size of 3 kb) and can be subdivided into a small number of families, GC-poor (L1 and L2), GC-rich (H1 and H2), and very GC-rich (H3). The GC-range of the isochores from the human genome is 30–60% (see also Fig. 2) (Modified from Bernardi, 1993a).

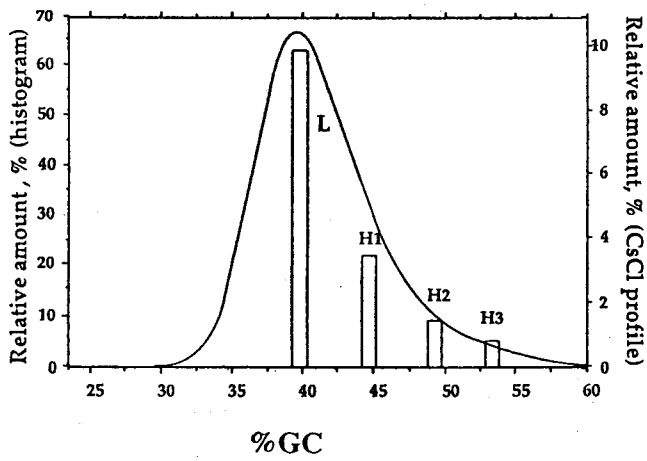

Fig. 2. Histogram of the isochore families from the human genome. The relative amounts of major DNA components derived from isochore families L (i.e., L1+ L2), H1, H2, H3 (see Saccone et al., 1993) are superimposed on the CsCl profile of human DNA. (Modified from Mouchiroud et al., 1991).

(Macaya et al., 1976), but they were so named (for 'equal regions') only later (Cuny et al., 1981).

The compositional distributions of large DNA fragments (namely plots of relative DNA amounts versus GC; see Fig. 3), of exons, of their codon positions (see Fig. 4 for the compositional distribution of third codon positions), and of introns represent the compositional patterns of genomes and define genome phenotypes (Bernardi and Bernardi, 1986). These are different in different vertebrate classes and even, to a smaller extent, within each vertebrate class (see Fig. 3), except for birds (see section **e**). Genome phenotypes are also different among plant genomes, to mention another group of organisms investigated in our laboratory (Salinas et al., 1988; Matassi et al., 1989; 1991; 1992; Montero et al., 1990). The concept of genome phenotype was introduced to stress the fact that the genome is not only a source of genetic information, but also a structure whose features, interactions and functions can be modulated by base composition (by changing codon usage, transcription rate, etc.).

(b) Compositional correlations and the genomic code

Linear compositional correlations exist (*i*) between exons (and their codon positions) and the isochores containing the corresponding genes (Fig. 5A), as well as between exons and introns of the same genes (Bernardi and Bernardi, 1985; 1986; Aïssani et al., 1991); and (*ii*) among codon positions (Bernardi and Bernardi, 1985; 1986; 1991; D'Onofrio and Bernardi, 1992; Fig. 5B). The former correlations (*i*) concern, in fact, coding sequences on the one hand and non-coding sequences on the other, and are different in the case of vertebrates and plants. In contrast, the second correlation (*ii*) is a universal correlation

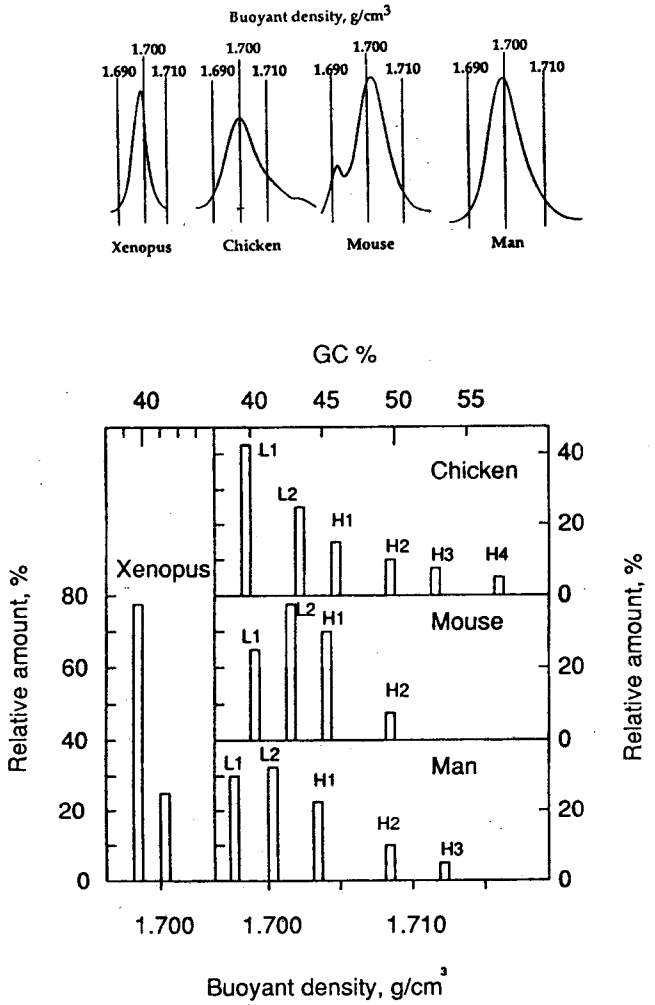

Fig. 3. Compositional patterns of vertebrate genomes. Top: CsCl profiles of DNAs from *Xenopus*, chicken, mouse and man. (From Thiery et al., 1976). Bottom: Histograms showing the relative amounts, modal buoyant densities and GC levels of the major DNA components from *Xenopus*, chicken, mouse, and man, as estimated after fractionation of DNA by preparative density gradient in the presence of a sequence-specific DNA ligand (Ag^+ or BAMD; BAMD is bis (acetato mercuri methyl) dioxane). The major DNA components are the families of large DNA fragments (see Fig. 1) derived from different isochore families. Satellite and minor DNA components (such as rDNA) are not shown in these histograms. (Modified from Bernardi, 1989).

valid from prokaryotes to man. These correlations indicate the existence of compositional constraints which act in the same direction, but to different extents, on both coding and non-coding sequences, as well as on the different codon positions. The correlations constitute a genomic code (Bernardi, 1990; 1993a), namely a set of rules which are obeyed by the nucleotide sequences forming the genomes.

(c) Gene distribution in isochores and chromosomal bands; compositional maps and gene density maps

The localization of genes in compositional fractions of DNA immediately led to the discovery that gene distribution is strikingly non-uniform in the human genome (Bernardi et al., 1985). The compositional correlations between third codon positions and the isochores containing the corresponding genes (Fig. 5A) can be used in order to investigate gene distribution (Mouchiroud et al., 1991; Bernardi, 1993a,b; S. Zoubak and G.B., in preparation). This approach has shown that gene concentration is low and constant over GC-poor isochores, increases over GC-rich isochores and reaches a maximum in the GC-richest isochores. This maximum is at least 20-times higher than the low, constant level found in GC-poor isochores (Fig. 6). The relative gene concentrations in the three compositional compartments of the human genome (GC-poor, GC-rich and very GC-rich) can be estimated as 4%, 20% and 76%, respectively (S. Zoubak and G.B., in preparation).

Because gene concentration parallels GC levels, a compositional map (Bernardi, 1989) of the human genome, or at least of regions of it, is of great interest because it is, in fact, a gene density map. Compositional maps can be constructed at the molecular (DNA) level or at the chromosomal level. In the first case, one can hybridize probes, that are located on a physical map, to fractionated DNA, and so determine the GC levels over 100–200 kb around the landmark corresponding to the probe (Bernardi, 1989; Gardiner et al., 1990; Pilia et al., 1993). In the second case, fractionated DNA is hybridized on metaphase chromosomes under conditions in which repeated sequences are competed out (Saccone et al., 1992; 1993). In some instances, a compositional map may be directly derived from sequence contigs (Ikemura and Aota, 1988; Ikemura et al., 1990) or, in the special case of yeast, from an entire chromosome sequence (Sharp and Lloyd, 1993; Karlin et al., 1993).

Compositional mapping at both DNA and chromosomal levels has shown (*i*) that G-bands are essentially formed by GC-poor isochores of the L1 and L2 families; (*ii*) that R′-bands (namely, R-bands exclusive of T-bands) are formed by GC-poor and GC-rich isochores (mostly of the H1 family) at comparable extents; and (*iii*) that T-bands (Dutrillaux, 1973; Ambros and Sumner, 1987) are formed by GC-rich isochores (mostly of the H2 family) and by the very GC-rich isochores of the H3 family. The latter point is in agreement with the finding that the genes which are richest in GC in their third codon positions are located in T-bands (Ikemura and Wada, 1991; De Sario et al., 1991). Fig. 7 presents the isochore distribution in chromosomal bands. On the basis of Figs. 6 and 7 and of other analyses, it was possible to estimate (S. Zoubak and G.B., in preparation) that 20% of genes are located in G-bands and 80% in R-bands. This is in agreement with previous values based on chromosomal localization of poly(A)$^+$RNA (Yunis

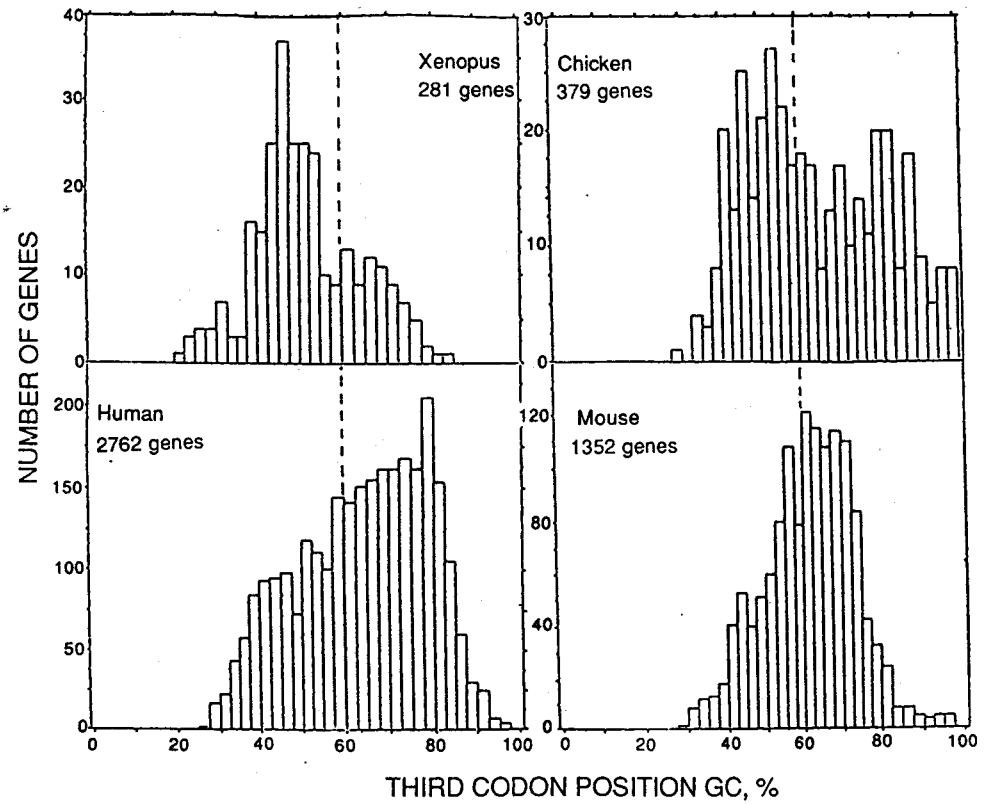

Fig. 4. Compositional distribution of third codon positions from vertebrate genes. The number of genes taken into account is indicated. The available gene sample of *Xenopus* was small, but the difference with the gene distribution from warm-blooded vertebrates was also found for homologous genes, indicating that this difference is not due to the sample used (see also Fig. 8). A 2.5% GC window was used. The broken line at 60% GC is shown to provide a reference (From Bernardi, 1993a).

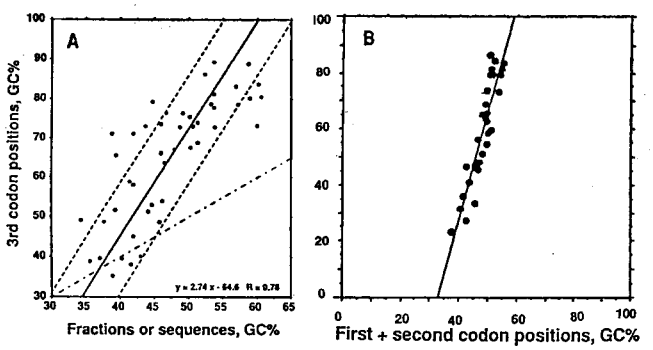

Fig. 5. Compositional correlations. (**A**) GC levels of third codon positions from human genes are plotted against the GC levels of DNA fractions (solid dots) or extended sequences (circles) in which the genes are located. The correlation coefficient and the slope are indicated. The dash-and-point line is the diagonal line (slope = 1). GC levels of third codon positions would fall on this line if they were identical to GC levels of surrounding DNA. The broken lines indicate a ±5% GC range around the slope (From Mouchiroud et al., 1991). (**B**) Plot of GC levels of third codon positions of genes from prokaryotic and eukaryotic genomes are plotted against GC levels of first + second positions. All values are averaged per genome (or per genome compartment, in the case of compositionally compartmentalized genomes) (From D'Onofrio and Bernardi, 1992).

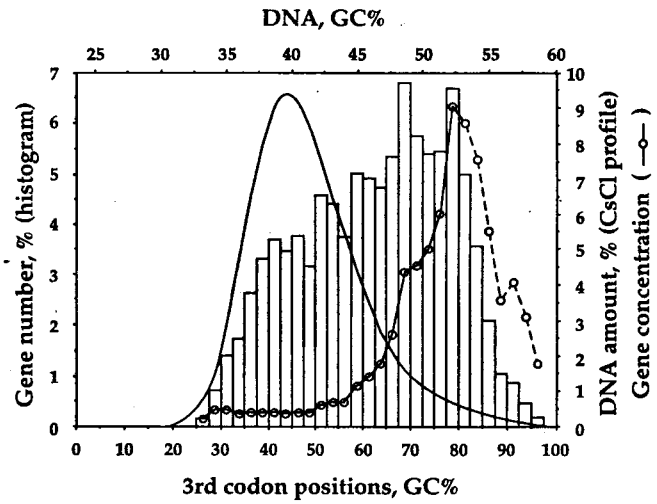

Fig. 6. Profile of gene concentration in the human genome as obtained by dividing the relative amounts of genes in each 2.5% GC interval of the histogram by the corresponding relative amounts of DNA, as deduced from the CsCl profile. It should be noted that the apparent decrease in gene concentration for very high GC values (broken line) is due to the presence of rDNA in that region; and that the last concentration values are uncertain because they correspond to very low amounts of DNA (From S. Zoubak and G.B., in preparation).

et al., 1978), and with values calculated for a different set of 1000 genes with known band location (Craig and Bickmore, 1993). However, the percentage of genes located in T-bands was estimated to be 58% by S. Zoubak and G.B. (in preparation), as against only 46% by Craig and Bickmore (1993).

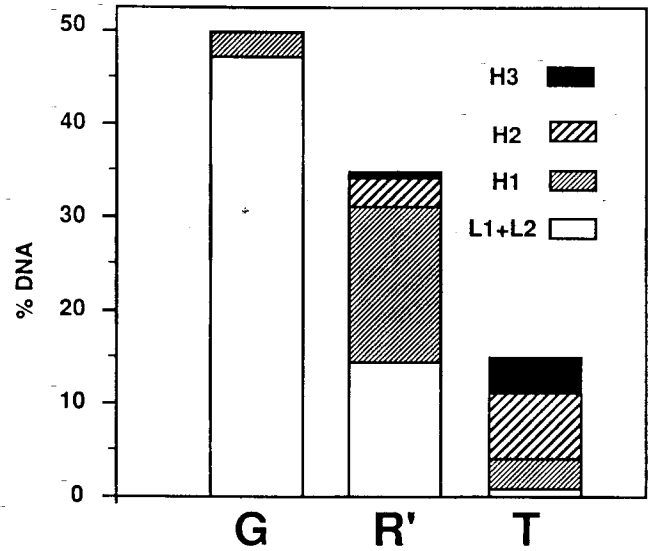

Fig. 7. A scheme of the relative amounts of isochore families L1+L2, H1, H2 and H3 in G-bands, R'-bands and T-bands. R'-bands are R-bands exclusive of T-bands (From Saccone et al., 1993).

(d) Some functional aspects of the isochore organization of the human genome

The gene concentration profile (Fig. 6) defines two very different sets of compositional compartments in the human genome. The first is characterized by a very low and constant gene density and corresponds to almost two thirds of the genome. The second one shows an increasingly higher gene density and corresponds to the remaining third or so of the genome. The latter compartments comprise the GC-rich isochores (mostly of the H1 family) of the R'-bands, and the GC-rich and very GC-rich isochores (of the H2 and H3 families) which essentially form the T-bands.

In the GC-rich compartments, the increasing gene concentration is paralleled by an increasing concentration of CpG islands (Aïssani and Bernardi, 1991a,b) and of CpG doublets (Bernardi et al., 1985; Bernardi, 1985). Since CpG islands are regularly associated with housekeeping genes but much less so with tissue-specific genes (Gardiner-Garden and Frommer, 1987), it is conceivable that housekeeping genes are predominantly located in the GC-rich compartments. In contrast with CpG islands and CpG doublets, *Alu* sequences do not parallel gene concentration, their peak coinciding with that of H2 isochores (Zerial et al., 1986; see Fig. 2).

The coding sequences of the GC-rich compartments increasingly stand out from the compositional background of intergenic sequences, in contrast to the coding sequences located in GC-poor isochores, which have the same GC level as the intergenic sequences (Aïssani et al., 1991). Moreover, the coding sequences of the GC-rich isochores are at least largely embedded in an open chromatin structure (Tazi and Bird, 1991; Aïssani and Bernardi, 1991a,b). Transcription and recombination reach the highest levels in the GC-rich and very GC-rich regions of the genome (see references in Bernardi, 1989; Rynditch et al., 1991; Eyre-Walker, 1993).

It should also be noted that R-bands replicate early in the cell cycle and condense late, whereas G-bands replicate late and condense early. Because of the gene distribution in isochores (see Fig. 6) and of the isochore distribution in R- and G-bands (see Fig. 7), both GC-rich and GC-poor genes should be expected to replicate early and late in the cell cycle, as it was indeed observed (Eyre-Walker, 1992). It is remarkable, however, that gene-rich chromosomal regions replicate early and gene-poor regions replicate late, and that the gene-richest regions are predominantly located at telomeres, namely at chromosomal sites that are tightly associated with the nuclear matrix and that interact with the nuclear envelope (de Lange, 1992).

Finally, as GC reaches the highest levels, both codon usage and amino-acid utilization become extreme (D'Onofrio et al., 1991). As far as the former is concerned, it should be recalled that at 100% GC in third codon position (a value approached by a number of the GC-richest human coding sequences; see Figs. 4 and 6) only 50% of the codons are used. Concerning aa utilization, those aa only comprising G and/or C in first and second codon positions, namely Gly, Ala, Pro and Arg (quartet), are abundant in proteins encoded by GC-rich coding sequences, whereas they are rare in proteins encoded by GC-poor coding sequences (D'Onofrio et al., 1991). In other words, genes located in GC-poor and GC-rich isochores greatly differ in codon usage and in the aa abundance of the encoded proteins.

A conclusion to be drawn from the results presented in this section is that, even if our knowledge on the subject is still limited, the isochore structure definitely has a number of important functional counterparts.

THE EVOLUTIONARY HISTORY OF THE ISOCHORE ORGANIZATION OF THE HUMAN GENOME

(e) The isochore patterns of vertebrate genomes

At this point, a crucial question should be asked, namely why a compositional map is a gene concentration map or, more specifically, why gene density, transcription, recombination, replication and condensation timing are correlated with GC levels of isochores in the human genome? This question can be answered, at least to a large extent, by comparing first the compositional patterns of the human genome with those of other mammals, of birds and of cold-blooded vertebrates, and then dis-

cussing the problem of gene distribution in vertebrate genomes (see section **f**).

(1) The mammalian patterns

The human genome pattern (see Figs. 2–4) is typical of the 'general' mammalian pattern which is by far the predominant genome pattern in mammals. This general pattern has been recognized in species belonging to 9 out of 10 orders investigated (Sabeur et al., 1993; Mouchiroud and Bernardi, 1993). Three 'special' patterns have been found in some mammalian genomes, one in pangolin (which belongs to the only genus of the order *Pholidota*), one in a megabat, and one in *Myomorpha* (megabats and *Myomorpha* belong to two orders, chiropters and rodents, respectively, in which the other sub-orders exhibit the general pattern). A comparison of the compositional patterns of man and mouse (Figs. 3 and 4) shows that the main difference between the general and the myomorph patterns is that the latter is narrower than the former (see also section below).

In the case of the general and of the myomorph patterns, homologous coding sequences from several species could be investigated. This showed the existence of a conservative mode of genome evolution (Bernardi et al., 1988), in which the base composition of codon positions are very remarkably conserved (Fig. 8). Indeed, not only first and second, but also third codon positions of homologous genes of mouse and rat (which exhibit the myomorph pattern) or of man and calf (which exhibit the general pattern) are very close in GC levels. This conservation is very striking if one considers two points: (*i*) that it applies not only to third codon positions having GC levels around 50%, but also to those having extreme values, over 90% GC (Bernardi et al., 1988, 1993); and (*ii*) that it accompanies large nucleotide divergences (Bernardi et al., 1988, 1993). In the case of the myomorph pattern, 12% of third codon positions are different in rat and mouse (which diverged about 20 Myr ago); in the case of the general pattern, 20% of third codon positions differ in their nucleotides in man and calf (which diverged at least 65 Myr ago). This second point suggests that mammalian genomes are at a compositional equilibrium, and that (because of the correlation of Fig. 5A) the hundreds of isochore pairs hosting the homologous genes investigated in the pairs of species under consideration are compositionally very close. Since the correlation of Fig. 5A has a general validity, all isochores containing homologous genes are compositionally close. This is in agreement with the existence of common compositional patterns at the DNA level, which were found experimentally (Sabeur et al., 1993). Moreover, if one considers the increasing evidence for the conservation of gene arrangements (linkage maps) among mammals (see O'Brien,

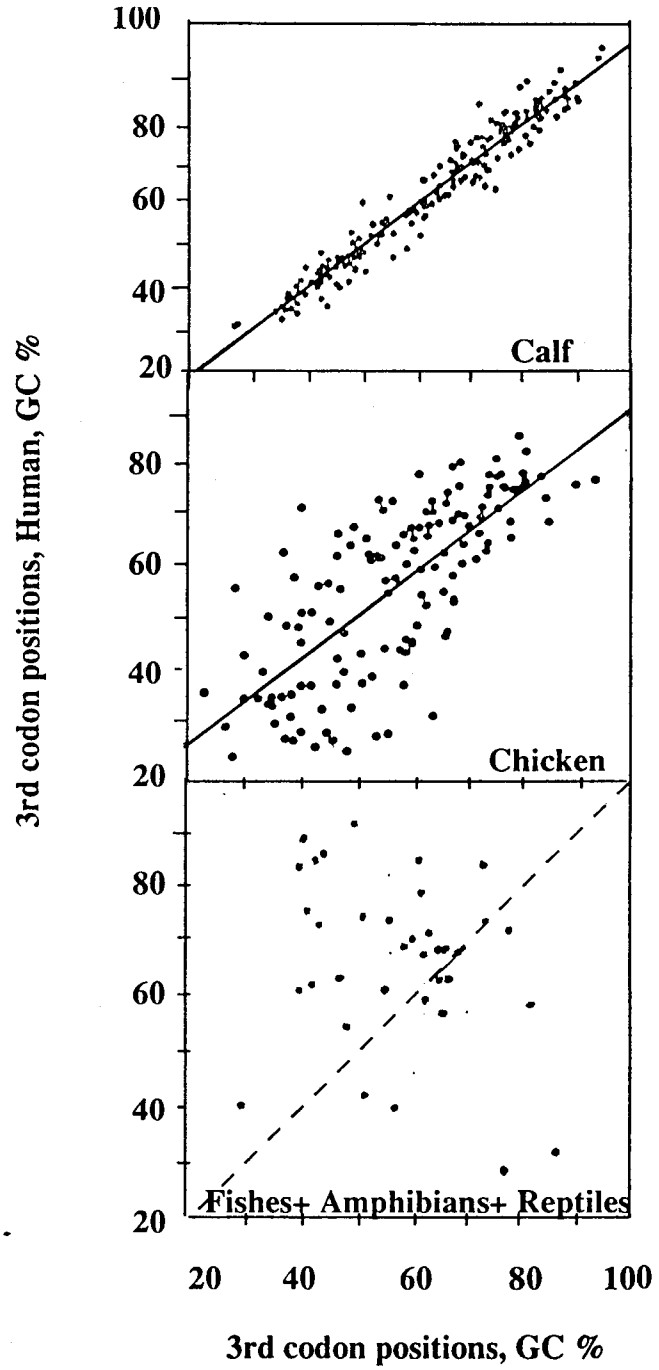

Fig. 8. GC levels of third codon positions of human genes (ordinate) are plotted agains those of their homologs from (top frame) bovine, (center frame) chicken and (bottom frame) cold-blooded vertebrates (abscissa). Least-square lines are shown in the top and center frames, the diagonal line (slope = 1) in the bottom frame (Modified from Mouchiroud and Bernardi, 1993; Kadi et al., 1993; and Bernardi and Bernardi, 1991; from top to bottom, respectively).

1993), one should draw the conclusion that, as a rule, syntenic regions, within either the general or the myomorph patterns, are compositionally similar.

It should be noted that, although the compositional pattern of myomorphs is narrower than the general pattern of mammals (see Figs. 3 and 4), the compositional

ranking order of third codon positions of homologous genes is very largely conserved (Mouchiroud et al., 1988; Mouchiroud and Bernardi, 1993). This indicates that the compositional 'compression' exhibited by the genome of myomorphs occurred in a most orderly way, as if by release of constraints on extreme values.

(2) The avian pattern

The conservative mode of evolution also appears to prevail in the case of birds, where a single compositional pattern was found, the avian pattern (Kadi et al., 1993). This mainly differs from those of mammals in that it reaches slightly higher GC levels (Figs. 3 and 4). A compositional comparison of codon positions from homologous genes of man and chicken (Fig. 8) revealed not only excellent correlations for second and first codon positions, but also a good one (the correlation coefficient being 0.78) for third codon positions (Fig. 8). This is a remarkable result for species which originated totally independently from each other, from different reptilian lineages and far apart (over 50 Myr) in time. The alternative explanation that the change already took place in the common ancestor of mammals and birds can be ruled out, because this ancestor, which existed more than 320 Myr ago (Carroll, 1987), was also at the origin of all reptiles (see Fig. 9), and none of present day reptiles investigated so far shows a compositional pattern similar to those of warm-blooded vertebrates (Bernardi and Bernardi, 1990a,b; 1991). Needless to say, that the third codon position correlations could also be explained by the unorthodox bird-mammal relationship proposed by Hedges et al. (1990). This might, however, be due to the compositional similarity of mammalian and avian genes used in constructing the phylogenetic tree (see Bernardi and Powers, 1992; Steel et al., 1993).

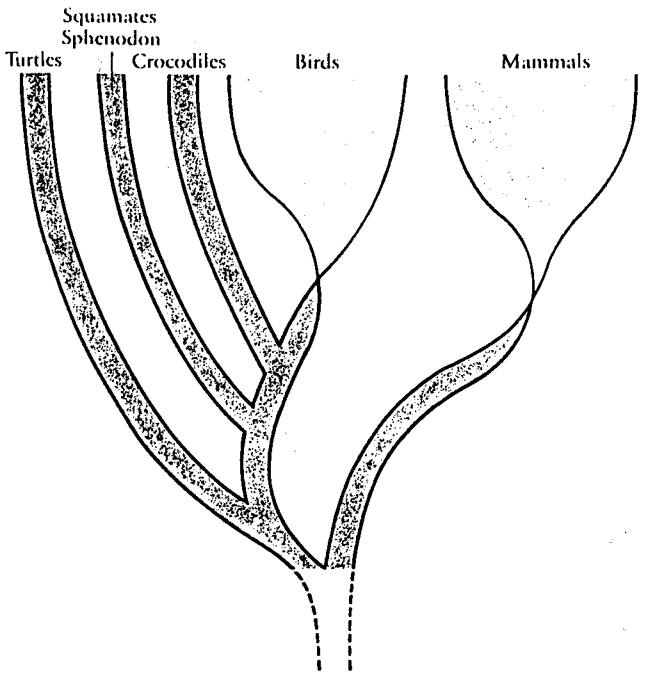

Fig. 9. Simplified phylogeny of amniotes. The classs *Reptilia*, as customarily defined, includes all the stippled lineages (From Carroll, 1987).

(3) The cold-blooded vertebrate patterns

In contrast with the results just mentioned for warm-blooded vertebrates, the comparison of human genes with homologous genes from cold-blooded vertebrates (Bernardi and Bernardi, 1991) shows that third codon positions of human genes are either equal or, more often, higher in GC than those of homologous genes from cold-blooded vertebrates; only exceptionally they are lower (Fig. 8). Similar differences can be observed at the DNA fragment (or isochore) level in that the human genome, as well as the genomes of warm-blooded vertebrates in general, attain higher GC levels and show a larger compositional heterogeneity than those of cold-blooded vertebrates (Bernardi and Bernardi, 1990a,b; 1991). Differences between cold- and warm-blooded vertebrates can also be seen at the chromosomal level. As suggested by Cuny et al. (1981) and then demonstrated by Medrano et al. (1988) and, in much more detail, by Schmid and Guttenbach (1988), chromosomes from cold-blooded vertebrates do not show any R-banding and only a poor G-banding or no G-banding, although they do show a replication banding (see references in Bernardi, 1989).

These differences exemplify the transitional or shifting mode in genome evolution (Bernardi et al., 1988), in which compositional changes take place in some regions of the genome (Figs. 10 and 11).

(f) Gene distribution in vertebrate genomes

The strikingly non-uniform distribution of genes in the genomes of warm-blooded vertebrates raises the question as to how genes are distributed in the compositionally uniform genomes of cold-blooded vertebrates. It was already suggested (Bernardi and Bernardi, 1990b) that all vertebrate genomes share a non-uniform and similar gene distribution. This suggestion was based on the general consideration that, if the gene distributions were different in cold- and warm-blooded vertebrates, this major evolutionary transition should have been accompanied by a most unlikely massive reshuffling of genes. In fact, it was recently demonstrated (S. Cacció, P. Perani, S. Saccone and G.B., in preparation) that, under conditions in which only single-copy sequence can hybridize, the GC-richest human isochores show homology not only with the GC-richest isochores of other mammals and birds, but also with the GC-richest isochores of cold-blooded vertebrates (the latter being obviously much lower in GC than the former). This indicates that the non-uniform gene dis-

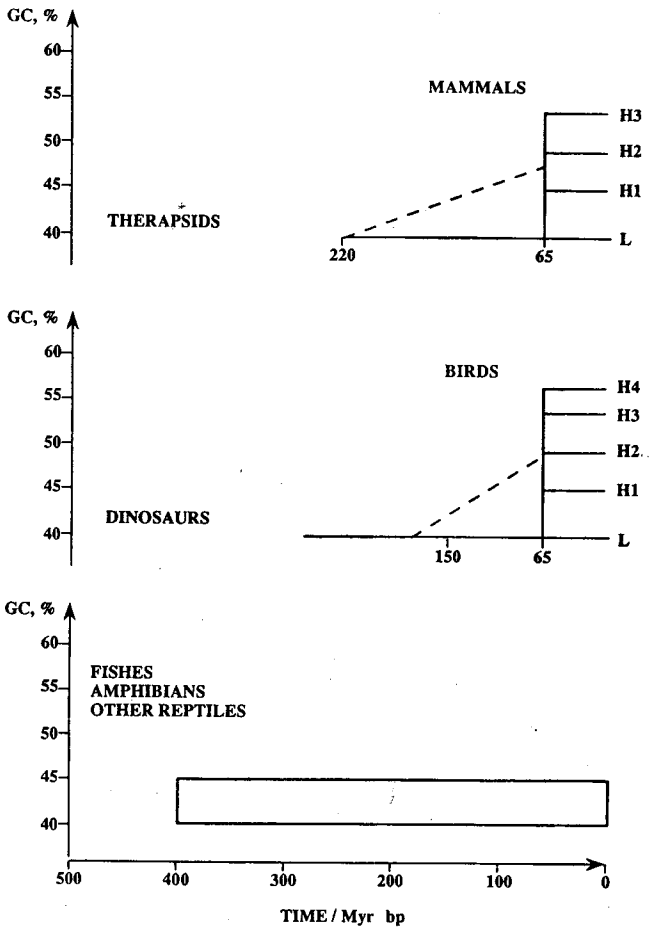

Fig. 10. Scheme of the formation of GC-rich isochores in the genome of warm-blooded vertebrates. While the genomes of cold-blooded vertebrates are characterized by a small range of compositional heterogeneity, those of mammals and birds underwent GC increases in some isochore families.

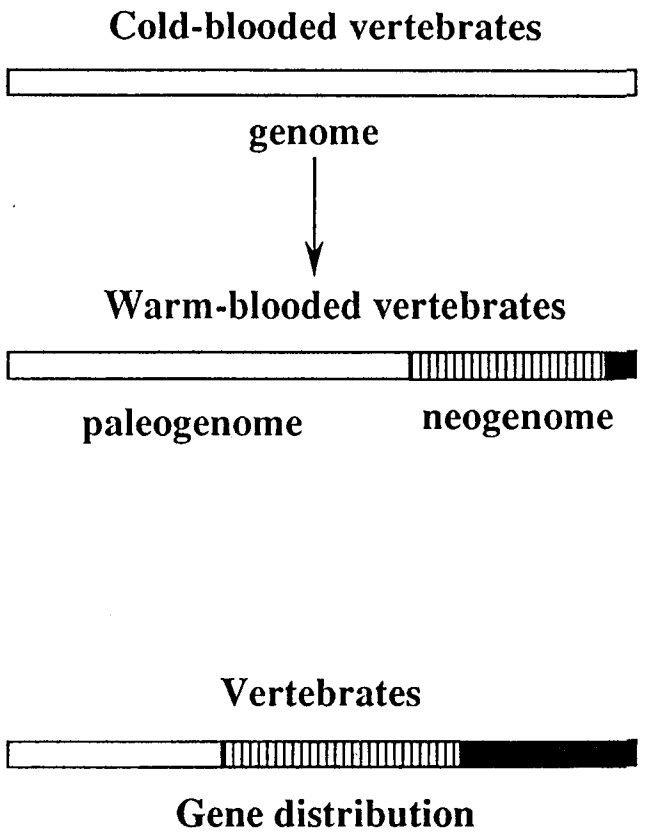

Fig. 11. Scheme of the compositional genome transition accompanying the emergence of warm-blooded from cold-blooded vertebrates. The compositionally homogeneous genomes of cold-blooded vertebrates are changed into the compositionally heterogeneous genomes of warm-blooded vertebrates. The latter comprise a paleogenome (corresponding to approximately two-thirds of the genome) that did not undergo any large compositional change and a neogenome (corresponding to the remaining one-third of the genome, with the GC-richest part only representing 4–5% of the genome). In the scheme, the mosaic structure of the warm-blooded vertebrate genome (see Fig. 1) is neglected; GC-poor isochores (open bar), GC-rich isochores (hatched bars), and very GC-rich isochores (black bar) are represented as three contiguous regions (From Bernardi, 1993c). The bottom bar represents gene concentration in the human genome (low, intermediate and high, respectively, in the GC-poor, GC-rich and GC-richest isochores). It indicates that gene distribution remains basically unchanged throughout vertebrates in spite of changes in genome size.

tribution found in the human genome is also present in cold-blooded vertebrates. In this connection one should recall here the existence of some, admittedly limited, syntenic regions shared by fishes and mammals (Morizot, 1990). It will be interesting to see whether the early/late replication pattern, which is shared by all vertebrates (see Bernardi, 1989, for references), also shows some correspondence between warm- and cold-blooded vertebrates, in that gene-rich regions might replicate early not only in warm- but also in cold-blooded vertebrates.

On the basis of the points mentioned above, the compositional transition which took place between reptiles and mammals or birds appears to have affected a pre-existing gene distribution pattern which was not basically modified by the compositional changes, in spite of the occurrence of genome size changes. The compositional transition concerned, by far and large, the same genes and their corresponding isochores in the temporally and phylogenetically independent events that led to the emergence of mammals and birds (see Figs. 8 and 10). A number of genes (like the globin genes) are known to be present on different isochores in mammals and birds (Bernardi et al., 1985), but, on the basis of Fig. 8, they should represent a minority of genes.

(g) The evolution of the vertebrate genome: the paleogenome and the neogenome

To sum up, the genomes of cold-blooded vertebrates are characterized by a low degree of compositional heterogeneity, their isochores covering, for each species, a narrow range of GC values. Still, gene distribution is highly non uniform, the highest GC levels within such narrow ranges being associated with the isochores that are the richest in GC; moreover, an early-late replication timing is present.

At two different geological times, the ancestors of present-day mammals and birds underwent similar GC increases in the gene-rich isochores of their genomes. Once the transitional mode of genome evolution had accomplished these changes, a conservative mode took over. This mantained within very narrow compositional levels the GC-rich isochores, including the extremely GC-rich third codon positions of a number of genes.

The part of the genome which was not changed compositionally at the reptile/warm-blooded vertebrate transition was called the paleogenome, the part which was changed the neogenome (Bernardi, 1989). Although only representing a minority, 35%, of the genome, the neogenome contains the majority of genes (Fig. 11), in man over 75% of them. It appears, therefore, that the isochores that underwent the compositional transition between reptiles and mammals or birds are characterized by a high gene concentration, and by a high transcription level (and, possibly, by early replication). This may even be a necessary condition for the compositional transition to take place, but certainly is not a sufficient one. Indeed, none of genomes from cold-blooded vertebrates investigated so far shows a compositional pattern like those of warm-blooded vertebrates.

The distinction between paleogenome and neogenome is of interest in another respect. Indeed, the GC-rich chromosomal bands from the compositionally heterogeneous genomes of warm-blooded vertebrates show a very high recombination rate (see references in Bernardi, 1989; Rynditch et al., 1991; Eyre-Walker, 1993) and these bands simply do not exist in cold-blooded vertebrates (Medrano et al., 1988; Schmid and Guttenbach, 1988). It has been suggested (Bernardi, 1993c), therefore, that the neogenome is responsible for the higher rate of karyotypic change and species formation of mammals and birds compared to the lower rates exhibited by cold-blooded vertebrates.

The results obtained on the compositional transition and on the compositional conservation of vertebrate genomes raise the question of the causes of such changes and such conservation. This point has already been discussed (Bernardi and Bernardi, 1986; Bernardi et al., 1988; Bernardi, 1993a) and will be dealt with further elsewhere (G.B., in preparation).

ACKNOWLEDGEMENTS

I wish to thank most heartily all my colleagues who participated in the investigations briefly reviewed in this paper. Their names and their contributions appear in the References. I wish to thank Giacomo Bernardi for useful discussions and Farida Kadi for her help in the preparation of some figures. The investigations reviewed in this paper received the support of AFM (Association Française contre les Myopathies), ARC (Association pour la Recherche contre le Cancer) and CNRS (Centre National pour la Recherche Scientifique).

REFERENCES

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

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

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

Ambros, P.F. and Sumner, A.T.: Metaphase bands of human chromosomes, and distinctive properties of telomeric regions. Cytogenet. Cell Genet. 44 (1987) 223–228.

Bernardi, G.: The organization of the vertebrate genome and the problem of the CpG shortage. In: Chemistry, Biochemistry and Biology of DNA Methylation (Cantoni, G.L. and Razin, A. (Eds.), Alan Liss, New York, 1985, pp. 3–10.

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

Bernardi, G.: Le génome des vertébrés: organisation, fonction, évolution. Biofutur 94 (1990) 43–46.

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

Bernardi, G.: The vertebrate genome: isochores and chromosomal bands. In: Sumner, A.T. and Chandley, A.C. (Eds.), Chromosomes Today, Chapman & Hall, London, 1993b, pp. 49–60.

Bernardi, G.: Genome organization and species formation in vertebrates. J. Mol. Evol. (1993c) in press.

Bernardi, G. and Bernardi, G.: Codon usage and genome composition. J. Mol. Evol. 22 (1985) 363–365.

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

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

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

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

Bernardi, G. and Powers, D.A.: Molecular phylogeny of the prickly shark, *Echinorhinus cookei*, based on a nuclear (18S RNA) and a mitochondrial (cytochrome b) gene. Mol. Phylogen. Evol. 1 (1992) 161–167.

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

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

Bettecken, T., Aïssani, B., Müller, C.R. and Bernardi, G.: Compositional mapping of the human dystrophin gene. Gene 122 (1992) 329–335.

Carroll, R.L.: Vertebrate Paleontology and Evolution. Freeman, New York, 1987.

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

Craig, J.M. and Bickmore, W.A.: Chromosome bands – flavours to savour. BioEssays 15 (1993) 349–354.

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

D'Onofrio, G., Mouchiroud, D., Aïssani, B., Gautier, C. and Bernardi, G.: Correlations between the compositional properties of human genes, codon usage and aminoacid composition of proteins. J. Mol. Evol. 32 (1991) 504–510.

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

de Lange, T.: Human telomeres are attatched to the nuclear matrix EMBO J. 11 (1992) 717–724.

Dutrillaux, B.: Nouveau système de marquage chromosomique. Chromosoma 41 (1973) 395–402.

Eyre-Walker, A.: Evidence that both G+C rich and G+C poor isochores are replicated early and late in the cell cycle. Nucleic Acids Res. 7 (1992) 1497–1501.

Eyre-Walker, A.: Recombination and mammalian genome evolution. Trans. Roy. Soc. B 252 (1993) 237–243

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

Gardiner, K., Aïssani, B. and Bernardi, G.: A compositional map of human chromosome 21. EMBO J. 9 (1990) 1853–1858.

Gardiner-Garden, M. and Frommer, M.: CpG islands in vertebrate genomes. J. Biol. Chem. 196 (1987) 261–282.

Hedges, S.B., Moberg, K.D. and Maxson, L.R.: Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of evidence for amniote relationship. Mol. Biol. Evol. 7 (1990) 607–633

Ikemura, T. and Aota, S.: Global variation in G+C content along vertebrate genome DNA. Possible correlation with chromosome band structures. J. Mol. Biol. 203 (1988) 1–13.

Ikemura, T. and Wada, K.: Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosomes numbers; relation between nucleotide sequence data and cytogenetic data. Nucleic Acids Res. 19 (1991) 4333–4339.

Ikemura, T., Wada, K. and Aota, S.: Giant G+C% mosaic structures of the human genome found by arrangement of GenBank human DNA sequences according to genetic positions. Genomics 8 (1990) 207–216.

Kadi, F., Mouchiroud, D., Sabeur, G. and Bernardi, G.: The compositional patterns of the avian genomes and their evolutionary implications. J. Mol. Evol. (1993) in press.

Karlin, S., Blaisdell, B.E., Sapolsky, R.J., Cardon, L. and Burge, C.: Assessments of DNA inhomogeneities in yeast chromosome, III. Nucleic Acids Res. 21 (1993) 703–711.

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

Matassi, G., Montero, L.M., Salinas, J. and Bernardi, G.: The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. Nucleic Acids Res. 17 (1989) 5273–5290.

Matassi, G., Melis, R., Macaya, G. and Bernardi, G.: Compositional bimodality of the nuclear genome of tobacco. Nucleic Acids Res. 19 (1991) 5561–5567.

Matassi, G., Melis, R., Kuo, K.C., Macaya, G., Gehrke, C.W. and Bernardi, G.: Large-scale methylation patterns in the nuclear genomes of plants. Gene 122 (1992) 239–245.

Medrano, L., Bernardi, G., Couturier, J., Dutrillaux, B. and Bernardi, G.: Chromosome banding and genome compartmentalization in fishes. Chromosoma 96 (1988) 188–183.

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

Morizot, D.C.: Use of fish gene maps to predict ancestral vertebrate genome organization. In: Isozymes: Structure, Function and Use in Biology and Medicine, Wiley-Liss, New York, 1990, pp. 207–234.

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

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

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

O'Brien, S.J. (Ed.) Genetic Maps, 6th Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993.

Pilia, G., Little, R.D., Aïssani, B., Bernardi, G. and Schlessinger, D.: Isochores and CpG islands in YAC contigs in human Xq26.1-qter. Genomics 17 (1993) 456–462.

Rynditch, A., Kadi, F., Geryk, J., Zoubak, S., Svoboda, J. and Bernardi, G.: The isopycnic, compartmentalized integration of Rous sarcoma virus sequences. Gene 10 (1991) 165–172.

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

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

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

Salinas, J., Matassi, G., Montero, L.M. and Bernardi, G.: Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. Nucleic Acids Res. 16 (1988) 4269–4285.

Schmid, M., and Guttenbach, M.: Evolutionary diversity of reverse (R) fluorescent chromosome bands in vertebrates. Chromosoma 97 (1988) 104–114.

Sharp, P. and Lloyd, A.T.: Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. Nucleic Acids Res. 21 (1993) 179–183.

Steel, M.A., Lockhart, P.J. and Penny, D.: Confidence in evolutionary trees from biological sequence data. Nature 364 (1993) 440–442.

Tazi, J. and Bird, A.: Alternative chromatin structure at CpG islands. Cell 60 (1991) 909–920.

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

Thomas, W.K. and Beckenbach, A.T.: Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. J. Mol. Evol. 29 (1989) 233–245.

Winkler, H.: Vererbung und Ursache der Parthenogenese im Pflanzen- und Tierreich, Fischer, Jena, 1920.

Yunis, J.J. and Tsai, M.Y.: Mapping of polysomal messenger RNA and heterogeneous nuclear RNA to the lightly staining G-bands of human chromosomes. Cytogenet. Cell Genet. 22 (1978) 364–367.

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