Evolutionary Genomics of Vertebrates and Its Implications

GIUSEPPE D'ONOFRIO,[a,b,e] KAMEL JABBARI,[a]
HÉCTOR MUSTO,[a,c,f] FERNANDO ALVAREZ-VALIN,[a,c]
STEPHANE CRUVEILLER,[a] AND GIORGIO BERNARDI[a,b,d]

[a]*Laboratoire de Génétique Moléculaire, Institut Jacques Monod 2, Place Jussieu, 75005 Paris, France*

[b]*Stazione Zoologica Anton Dohrn, Villa Comunale 1, 80121 Naples, Italy*

[c]*Facultad de Ciencias, Iguá 4225 Montevideo 11200, Uruguay*

ABSTRACT: The discovery that the vertebrate genomes of warm-blooded vertebrates are mosaics of isochores, long DNA segments homogeneous in base composition, yet belonging to families covering a broad spectrum of GC levels, has led to two major observations. The first is that gene density is strikingly non-uniform in the genome of all vertebrates, gene concentration increasing with increasing GC levels. (Although the genomes of cold-blooded vertebrates are characterized by smaller compositional heterogeneities than those of warm-blooded vertebrates and high GC levels are not attained, their gene distribution is basically similar to that of warm-blooded vertebrates.) The second observation is that the GC-richest and gene-richest isochores underwent a compositional transition (characterized by a strong increase in GC level) between cold- and warm-blooded vertebrates. Evidence to be discussed favors the idea that this compositional transition and the ensuing highly heterogeneous compositional pattern was due to, and was maintained by, natural selection.

Thirty years ago we found that DNA-silver complexes could be fractionated in Cs_2SO_4 density gradients according to the frequency of silver-binding sites on DNA molecules, thus allowing high-resolution, sequence-dependent fractionation of DNA.[1] This approach led to the discovery of the striking **compositional heterogeneity** of high molecular weight, "main band" (nonsatellite) bovine DNA[2] and to the subsequent findings that (1) vertebrate genomes are mosaics of **isochores** (FIG. 1), namely, long DNA segments (>300 kb), which are compositionally homogeneous (above a size of 3 kb) and belong to a small number of families characterized by different GC levels (GC is the molar fraction of guanine+cytosine in DNA); (2) **isochore families** define the **isochore pattern** of a genome, which can be investigated on the large DNA fragments (approximately 100 kb in size) that make up routine DNA preparations; and (3) isochore patterns differ between cold- and warm-blooded vertebrates.[3,4]

This straightforward approach, based on the nucleotide composition of DNA sequences, not only has proven to be very powerful for the study of genome organization, especially in the case of complex eukaryotic genomes, but also has led to novel insights with interesting evolutionary implications. In this brief review (see also ref. 5) we first present a summary of our current knowledge of the sequence organization of the human genome, which is typical of mammalian genomes (these, in turn, share their basic properties with avian genomes). We then move into the evolutionary implications of our findings.

[d]Phone, +33–1 44 27 81 72; fax, +33–1 44 27 79; e-mail, bernardi@citi2.fr
[e]Phone, (00)39 81 5833296; fax, (00)39 81 7641355; e-mail, donofrio@alpha.szn.it
[f]Phone, (00) 598 2 48 95 31; fax, (00) 598 2 40 99 73; e-mail, hmusto@genetica.edu.uy

The **compositional pattern** of the human genome, at the DNA level, is characterized (FIG. 1) by GC-poor, L, isochores that represent about 63% of the genome, whereas GC-rich, 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.[6] Another type of compositional pattern is that of coding sequences; in this case, either GC

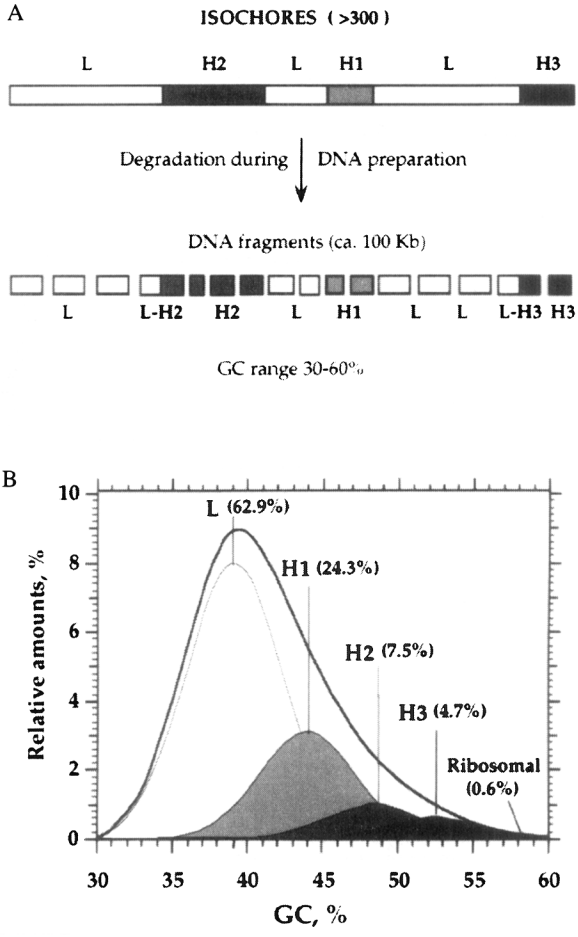

FIGURE 1. **(A)** Scheme of isochore organization of the human genome. This genome, which is typical of the genome of most mammals, is a mosaic of large (>300 kb on average) DNA regions, the isochores, which are compositionally homogeneous (above a size of 3 kb) and can be partitioned into a smaller number of families, "light" or GC poor (L1 and L2; here collectively indicated as L), "heavy" or GC rich (H1 and H2), and very GC rich (H3). Isochores are degraded during routine DNA preparation to fragments of approximately 100 kb in size. The GC range of isochores from the human genome is 30–60%. **(B)** 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, and H3. The relative amounts of major DNA components and their modal GC values are indicated. Satellite DNAs are not represented.

or, more informatively, GC_3 (the GC level of third codon positions) defines the pattern (see, for example, the human GC_3 pattern in FIG. 3). Yet an alternative pattern is represented by the compositional distribution of introns, which, however, is less useful because intron sequences, so far, are poorly represented in data banks. Compositional patterns have been called **genome phenotypes**,[7] because they differ not only between cold- and warm-blooded vertebrates, as already mentioned, but also, to a lesser extent, among different vertebrate orders within a class and even among different families within an order.[8–10]

An obvious question is whether any correlation exists between the compositional patterns of coding sequences (which may represent as little as 3% of the genome in vertebrates) and the compositional patterns of DNA fragments (which are formed by intergenic sequences and introns and may represent 97% of the genome). Another question is whether there is any correlation between the compositions of the coding sequences and the introns from the same genes. The answer to both questions is yes. Indeed, linear correlations hold between GC levels (in particular, GC_3 levels) of coding sequences and the GC levels of the isochores in which coding sequences are embedded; likewise, linear correlations hold between GC levels of coding sequences and GC levels of the corresponding introns.[11,12] These **genome equations** (FIG. 2), together with the equation of the universal correlations to be discussed, amount to a **genomic code**.[13]

The correlation between GC_3 of coding sequences and GC of isochores (FIG. 2) is especially important because it allows positioning of the distribution profile of coding sequences relative to that of DNA fragments. In turn, this allows calculating the gene density by dividing the percentage of genes located in given GC intervals by the percentage of DNA located in the same intervals. 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. 3), gene concentration increasing from very low levels in L isochores to 20-fold higher levels in H3 isochores.[6,11,14]

H3 isochores have been called the **genome core**,[13] because of their very high gene concentration (1 gene per 5–10 kb), which is comparable to those of compact eukaryotic genomes, and because of their functional properties (FIG. 4). Indeed, genes located in H3 are mostly associated with CpG islands,[15–17] comprise most or all housekeeping genes, are actively transcribed, and are characterized by an open chromatin structure[18] which features scarcity, or absence, of H1 histones, acetylation of H3 and H4 histones, and a large nucleosome spacing.[19] H3 isochores show their highest concentrations in the most thermal-denaturation–resistant reverse bands of metaphase chromosomes, the $H3^+$ bands,[20–22] which undergo very active recombination and replicate earliest in the cell cycle.[23] In contrast, GC-poor isochores exhibit a generally closed chromatin structure, with only a few open regions corresponding to genes that are largely transcribed in a tissue-specific or developmentally regulated manner (FIG. 4) and are located in Giemsa bands and $H3^-$ reverse bands, in which H3 isochores cannot be detected.

The results just outlined raise the question, what is the evolutionary background of the compositional properties and of the gene distribution of the human genome?

Compositional analysis of the genomes of all vertebrate classes shows that the compositional pattern just described for the human genome is basically shared by all warm-blooded vertebrates. 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 very general rule (see, however, Bernardi *et al.*[24]), they do not reach, by far, the high GC levels attained by the genomes of warm-blooded vertebrates.[3,25–27] Genes, however, are

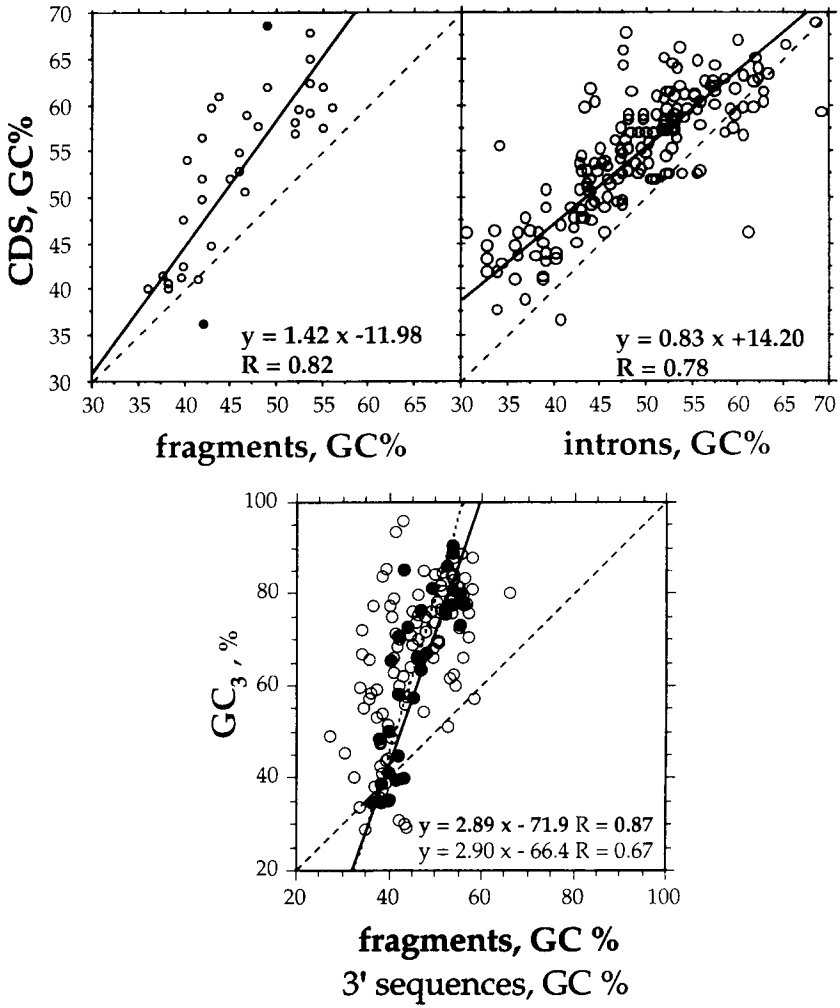

FIGURE 2. Correlation between GC of coding sequences and the GC levels of the large DNA fragments in which sequences were localized, or the GC levels of the corresponding introns (*top frames*); *black dots* were not considered in the calculation of the regression line. The *bottom frame* shows the correlation between GC_3 of coding sequences (*filled circles*) and of 3′ flanking sequences further than 500 bp from the stop codon (*open circles*) and the GC levels of the DNA fractions in which the genes were localized. *Solid* and *dotted lines* are the regression lines through the two sets of points, respectively. The unity slope lines (*broken diagonal lines*) are also shown.

not uniformly distributed in these genomes either. Indeed, only the GC-richest 10% or so of the compositional DNA fractions of cold-blooded vertebrates hybridize single-copy DNA from human H3 isochores.[5] This indicates that in cold-blooded vertebrates as well, genes are located in the GC-richest fractions of the genomes, even if these GC-richest

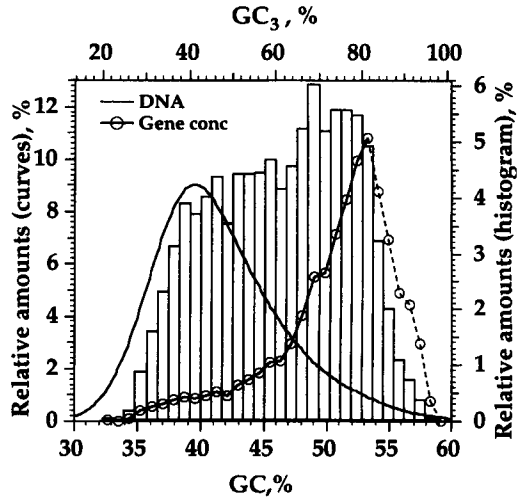

FIGURE 3. Profile of gene concentration (*circles*) in the human genome as obtained by dividing the relative numbers of genes in each 2% GC interval of the gene distribution histogram by the corresponding relative amounts of DNA deduced from the CsCl profile. Positioning of the GC_3 histogram relative to the CsCl profile is based on the correlation of FIGURE 2 (*bottom frame*). 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.

fractions are much less GC-rich than the corresponding fractions of warm-blooded vertebrates. These findings indicate that a compositional genome transition took place between cold- and warm-blooded vertebrates and that this transition concerned a small part of the genome, interestingly the gene-richest part of it.

The best evidence of the compositional transition is provided by comparisons of GC_3 values of orthologous genes (FIG. 5). When such GC_3 plots concern genes from human and other mammals sharing the "general mammalian pattern," such as calf, the regression line goes 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

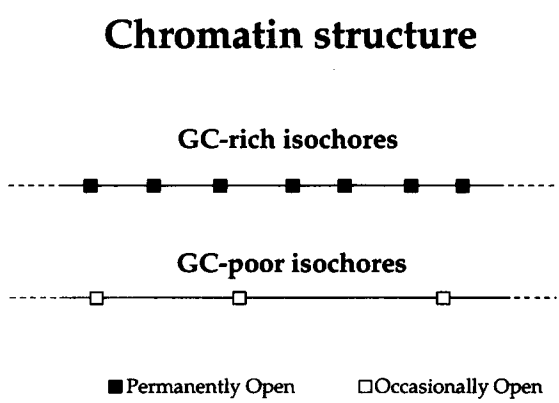

FIGURE 4. Scheme of transcribing regions in GC-rich and GC-poor isochores. In the former, genes are very close to each other and, to a large extent, are transcribed constitutively (*solid boxes*). In the second, genes are very sparse and are largely transcribed in a tissue-dependent or developmentally regulated manner (*open boxes*). Note that the actual ratio of gene concentrations in H3 and L isochores is close to 20:1.[6]

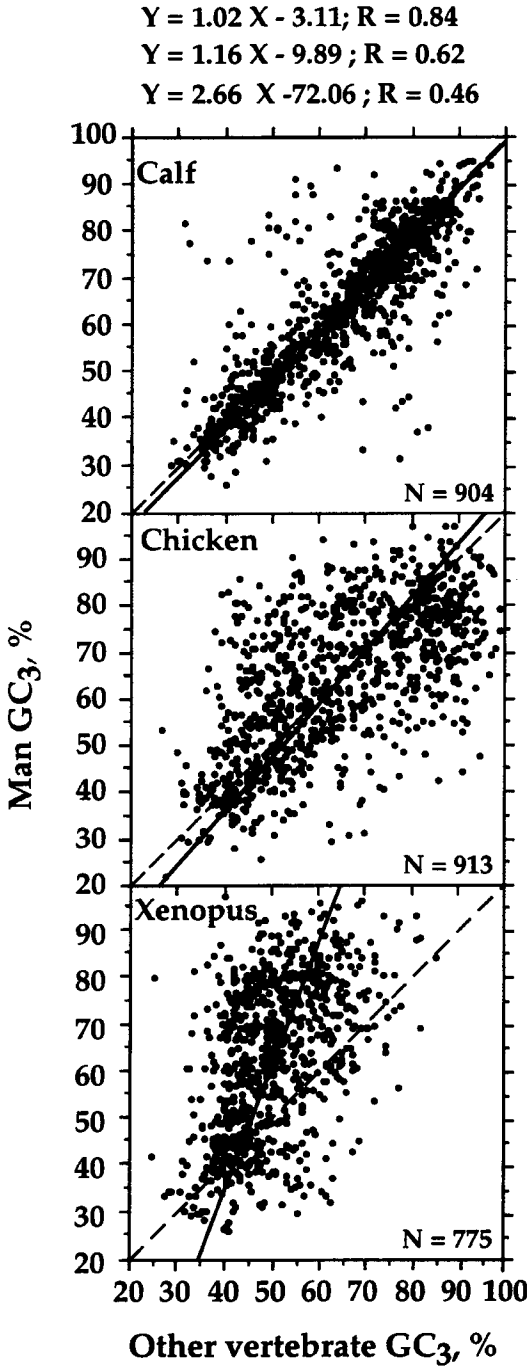

FIGURE 5. Correlations between GC_3 values of orthologous genes from human and calf, human and chicken, and human and *Xenopus*. N is the number of gene pairs analyzed. Orthogonal regression lines (*solid lines*) are shown together with the diagonals (*broken lines*). Corresponding equations are shown at top of figure.

close to each other. When the plot compares GC_3 values of orthologous genes from human and chicken, the slope is close to unity, the straight line practically coincides with the diagonal, but the correlation coefficient is lower (due to the larger scatter of points), yet still highly significant. This stresses three important points, namely, that (1) mostly the same genes were affected by the compositional transition in the two independent lines of mammals and birds; (2) these genes have a similar isochore distribution in mammals and birds; and (3) this similar distribution reflects that of their common ancestral reptile. In contrast, when the plot concerns human and *Xenopus* genes, points are scattered around the diagonal in the low GC range, showing little compositional change between the two species, but human values are increasingly higher, on the average, than the corresponding *Xenopus* values, as increasingly higher GC ranges are explored. As a result, the slope, 2.7, is much higher than unity.

At this point, we note that the compositional transition just described (1) concerned only the GC-rich genes from the gene-dense regions, namely, the genome core; (2) occurred (and was similar) in the independent ancestral lines of mammals and birds, but in no cold-blooded vertebrate; and (3) stopped with the appearance of present-day mammals and birds, as indicated by the essentially identical patterns found in different mammalian orders (such as primates and artiodactyls; see FIG. 6) that diverged from each other some 100 million years ago, as well as in different avian orders.[9,24] This means that the convergent compositional evolution, undergone by the genome core of the ancestors of mammals and birds, apparently reached a compositional equilibrium at the time of appearance of mammals and birds, and that from that time on the compositional changes associated with the cold- to warm-blooded transition were maintained to the present day. This is remarkable, if we consider that 38% coding sequences from warm-blooded vertebrates are characterized by GC_3 levels higher than 70%. Incidentally, these high GC_3 levels cause a strong bias in codon usage; indeed, at 100% GC_3, only 50% of codons are available.

Two obvious questions arise about the cause(s) of (1) the compositional genome transitions of vertebrates and (2) the maintenance of the new compositional patterns. The original explanation for the compositional transition[7] was that natural selection was responsible for it. The selective advantages for the vertebrates that are characterized by high body temperatures were considered to be the higher thermal stability of DNA, of RNA, and of the important proteins encoded by the GC-rich coding sequences (e.g., housekeeping genes), all these advantages simultaneously being achieved by the compositional transition.

As for the first two points, recall that H3 isochores are located[20,22] in reverse chromosomal bands that had previously been identified as bands that are particularly resistant to thermal denaturation[28] and that abundant evidence exists in favor of the fact that GC increases stabilize RNA structures (see, for example, refs. 29–31).

The third point requires a more detailed discussion. We recall that universal correlations were reported[32,33] between the GC levels of the three codon positions which hold both intragenomically (within the coding sequences of the human genome, for example) and intergenomically (within the coding sequences of different genomes, as averaged per genome or per isochore family in the case of compartmentalized genomes). These correlations have been reanalyzed using a vastly enlarged data set, and the results obtained are shown in FIGURE 6. An implication of the positive correlations between GC_3 and GC_1 or between GC_3 and GC_2 is that GC_3 increases should be accompanied by (1) 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, this was found to be the case (not

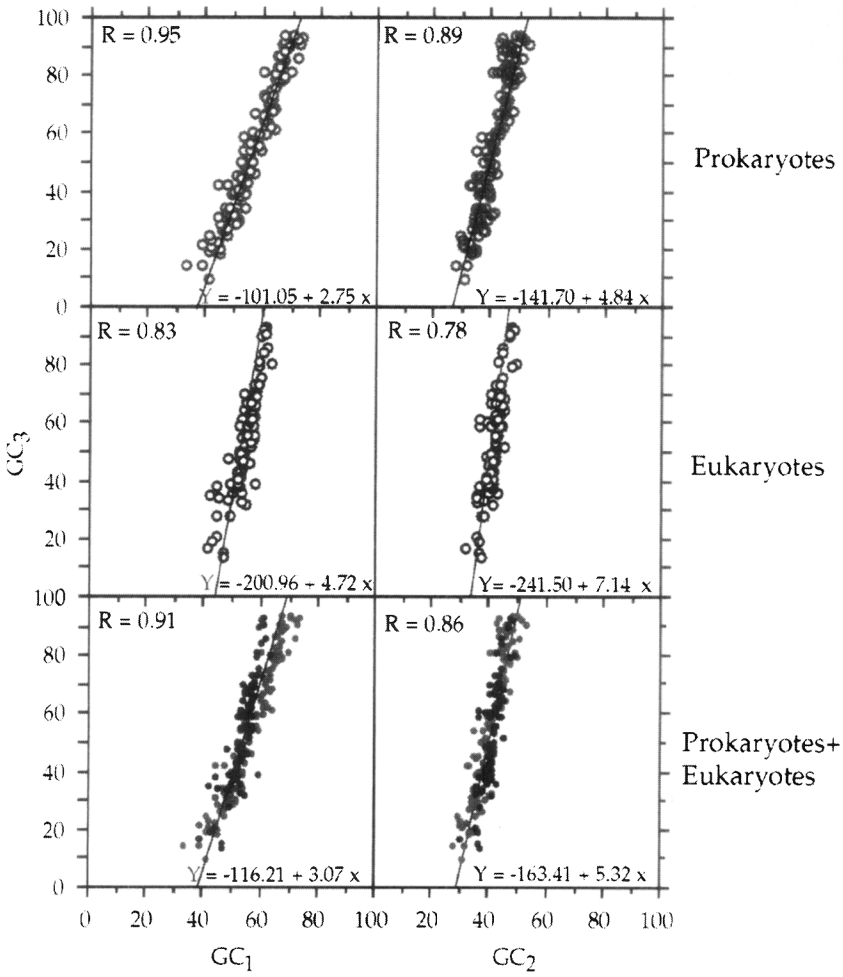

FIGURE 6. Intergenomic compositional correlations. Plots for prokaryotes, eukaryotes, and prokaryotes + eukaryotes are shown, along with the equations of orthogonal regression lines and correlation coefficients. In the case of heterogeneous eukaryotic genomes, GC_3 values of genes averaged per genome compartments (GC-poor, GC-intermediate, and GC-rich) are plotted against the corresponding GC_1 and GC_2 values.

shown; see, however, FIG. 7, which displays the Grantham representation of the genetic code); and (2) increases of amino acids encoded by codons of the GC class (namely, those codons made up of G and/or C in first and second positions[32] and decreases of amino acids encoded by codons of the AT class (namely, those codons made up of A and/or T in first and second positions). This was verified[32] and rechecked on recent data (not shown).

A more recent finding is that GC_3 is positively correlated not only with the frequency of amino acids of the GC class, but also with the hydropathy of the amino acids, from bac-

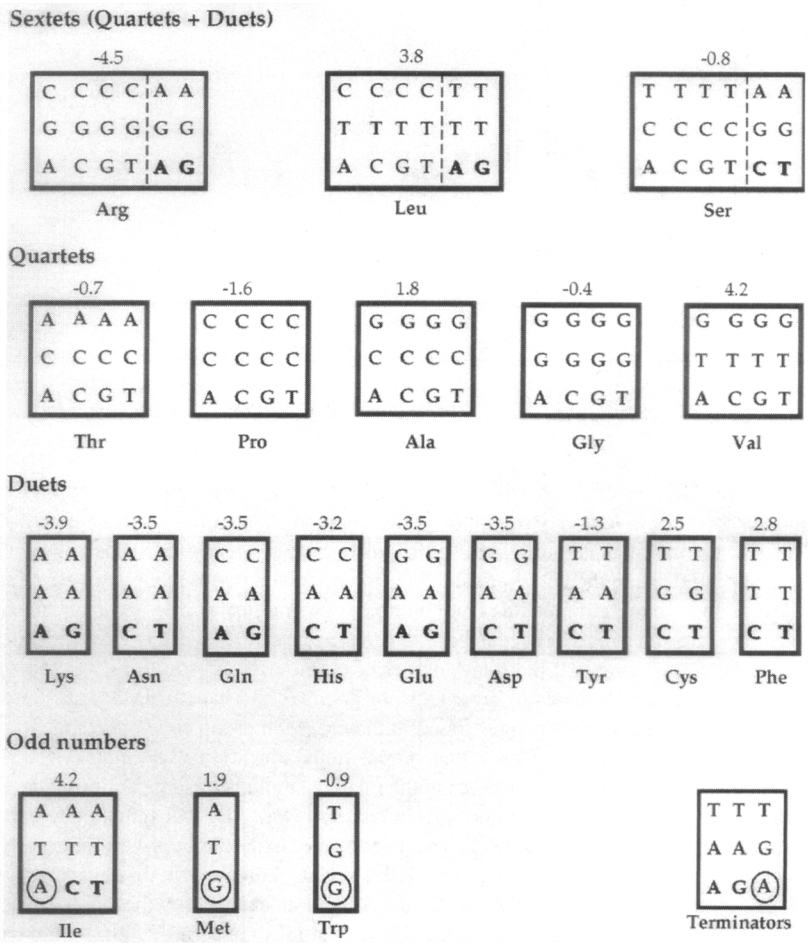

FIGURE 7. The genetic code. The Grantham[43] representation was modified in that codons rather than anticodons are shown, a distinction is made among third position nucleotides of quartet, duet, and odd number codons, and hydropathy values for amino acids[44] are shown.

teria and vertebrates (FIG. 8). Interestingly, although the slopes of the two regression lines were identical, prokaryotic values were systematically more hydrophobic than vertebrate values. This difference was accompanied by another remarkable property of prokaryotic versus eukaryotic proteins, namely, that the former had a cysteine level half as high as that of vertebrates. These novel findings stress the existence of unexpected global differences between proteins from prokaryotes and those from vertebrates.

At this point, we pushed further our analysis of homologous genes from *Xenopus* and man. Genes were divided into three groups, according to GC_3 levels of human coding sequences: 0–45%, 45–65%, and 65–100%. When amino acid compositions of the first

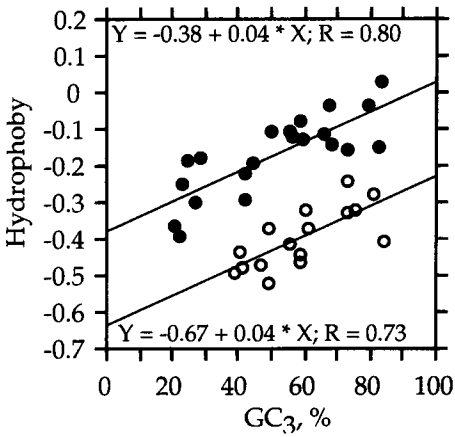

FIGURE 8. Correlation between hydropathy values (using the Kyte and Doolittle scale[44]) and GC_3 for amino acids pooled from individual prokaryotic (*closed circles*) and vertebrate genomes (*open circles*).

and last groups were compared (FIG. 9A), differences were negligible in the GC_3-poor group, whereas they were very pronounced for about half the amino acids of the GC_3-rich group. (In the intermediate group, not shown, differences were close to those of the GC_3-poor group.) Indeed, in the GC_3-poor and GC_3-intermediate groups, changes were as small as those found among amino acids from homologous proteins of calf and man (maximal values covering the range between the two broken lines of FIG. 9A). These results are remarkable, considering that the total number of amino acid substitutions was the same, about 20% in the three classes of proteins from *Xenopus* and man. They are also remarkable because the directional changes found in the GC_3-rich group concern as much as one fourth of all changes. In other words, the compositional transition affecting GC-rich genes was accompanied by important changes in the relative amounts of some amino acids. Consideration of the hydropathy of amino acids revealed (FIG. 9B) that remarkable changes took place in the GC_3-rich group leading to a net gain of hydrophoby, whereas, expectedly, they did not occur in the GC_3-poor group. These results indicate that, on the average, proteins encoded by GC_3-rich genes underwent structural changes at the cold- to warm-blooded transition (FIGS. 9 and 10). Current investigations concern the different extent of changes in different proteins and the localization of changes within the polypeptide chains.

Very different behavior of the GC-rich compared to GC-poor genes was found not only at the compositional transition taking place between cold- and warm-blooded vertebrates, but also in the maintenance of the directional changes in mammalian genes. Indeed, when GC-rich homologous coding sequences from four orders of mammals (which were separated for about 100 millions years) were investigated, it was found that (1) the frequencies of synonymous substitutions of quartet (fourfold degenerate) codons were gene specific, correlated with the frequencies of non-synonymous substitutions and significantly deviated from expectations based on a stochastic process in which nucleotide substitutions accumulate at random over time (this being the case, in contrast, for GC-poor coding sequences[34,35]); (2) synonymous positions (especially conserved positions) of quartet codons exhibited significantly different base compositions compared to expectations based on a "random" substitution process from the "ancestral" (consensus) sequence to the present day (actual) sequences, whereas significant differences were rare in GC-poor genes;[36] and (3) intragenic variability of synonymous rates was correlated with that of

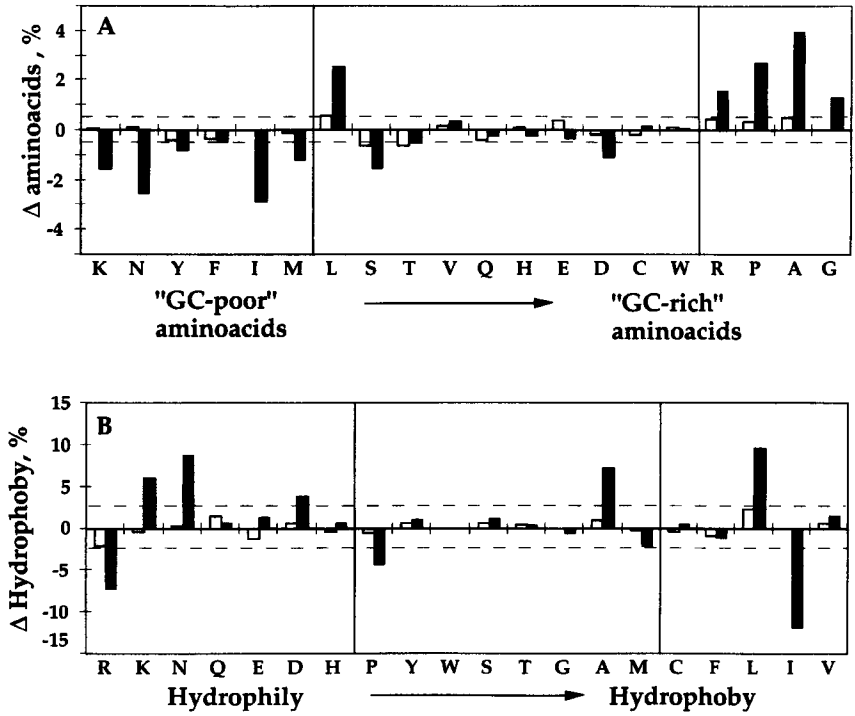

FIGURE 9. Difference histogram of relative amounts (**A**) and hydrophoby (**B**) of amino acids from homologous proteins of man and *Xenopus*. Proteins were partitioned into three groups according to the GC_3 values of the corresponding human coding sequences. The GC_3-rich group (*closed boxes*) and the GC_3-poor group (*open boxes*) are shown. The *horizontal broken lines* give the maximal difference values for calf and human homologous proteins. The *vertical lines* delimit the GC-poor and GC-rich classes of amino acids (in **A**) and hydrophobic and hydrophilic amino acids (in **B**).

non-synonymous rates; moreover, the variation in GC level (and especially in C level) of all silent positions along each gene was correlated with the variation in synonymous rate.[37] These results indicate that synonymous and non-synonymous rates as well as GC levels of synonymous positions of GC-rich coding sequences are under some common selective constraints, the constraints in synonymous positions being possibly related to selection for translational accuracy.[38]

Now it should be mentioned that an alternative interpretation for the regional GC increases undergone by the genome core of warm-blooded vertebrates is that a mutational bias (see, for example, Li[39]) was responsible for it. This interpretation, however, is problematic in several ways. Indeed, it is difficult to understand (1) why the regional compositional changes leading to the formation of GC-rich isochores never occurred in any cold-blooded vertebrate; (2) why they only occurred in a very small part of the vertebrate genome; (3) why they paralleled each other in mammals and birds; (4) why they reached

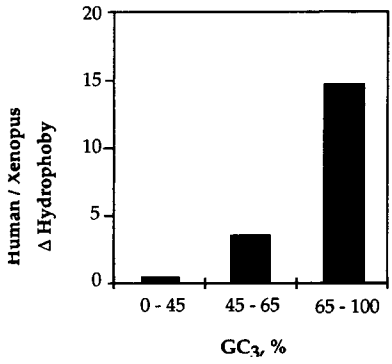

FIGURE 10. Difference histogram of hydrophoby of amino acids from homologous proteins of man and *Xenopus*. Proteins were partitioned into three groups according to the GC_3 values of the corresponding human coding sequences.

an equilibrium a long time ago, in both mammals and birds, and were maintained since; and (5) why they were different in different regions of the same gene.

It should also be stressed that the low GC levels of some thermophylic bacteria do not contradict, as claimed,[31] the selectionist interpretation given above. Indeed, different strategies were developed by organisms to cope with long-term high body temperatures. It is known that the DNAs of such bacteria are stabilized by particular DNA-binding proteins[40] and that their proteins can be stabilized by thermostable chaperonins.[41]

In conclusion, recent results from our laboratory support the original working hypothesis that selective advantages underlie the regional compositional changes accompanying the transition from cold- to warm-blooded vertebrates;[7] moreover, they maintain the novel, high GC levels attained.[34,36] They considerably refine the original idea in stressing that (1) selection concerns a small, yet functionally the most important part of the vertebrate genome, the majority of which is not under comparable constraints; (2) two distinct classes of coding sequences can therefore be distinguished in the vertebrate genome, only one of which is under evident selection; (3) distinct regions can be detected inside coding sequences, which have different properties, in terms of synonymous and non-synonymous mutation rates; (4) GC increases in intergenic and intragenic (intronic) non-coding sequences of the genome core can be explained by the fact that these sequences, which are very short in the genome core,[6,42] are largely endowed with regulatory roles, and selection can therefore operate on them as well; and (5) repair being much more active in transcribed compared to nontranscribed or poorly transcribed sequences, the selection load is very much reduced in the genome core and may be compatible with the small population size of vertebrates.

REFERENCES

1. Corneo, G., E. Ginelli, C. Soave & G. Bernardi. 1968. Isolation and characterization of mouse and guinea pig satellite DNA's. Biochemistry **7:** 4373–4379.
2. Filipski, J., J.-P. Thiery & G. Bernardi. 1973. An analysis of the bovine genome by Cs_2SO4-Ag+ density gradient centrifugation. J. Mol. Biol. **80:** 177–197.
3. Thiery, J.P., G. Macaya & G. Bernardi. 1976. An analysis of eukaryotic genomes by density gradient centrifugation. J. Mol. Biol. **108:** 219–235.
4. Macaya, G., J.-P. Thiery & G. Bernardi. 1976. An approach to the organization of eukaryotic genomes at a macromolecular level. J. Mol. Biol. **108:** 237–254.

5. BERNARDI, G. 1995. The human genome: Organization and evolutionary history. Annu. Rev. Genet. **29:** 445–476.
6. ZOUBAK, S., O. CLAY & G. BERNARDI. 1996. The gene distribution of the human genome. Gene **174:** 95–102.
7. BERNARDI, G. & G. BERNARDI. 1986. Compositional constraints and genome evolution. J. Mol. Evol. **24:** 1–11.
8. SABEUR, G., G. MACAYA, F. KADI & G. BERNARDI. 1993. The isochore patterns of mammalian genomes and their phylogenetic implications. J. Mol. Evol. **37:** 93–108.
9. KADI, F., D. MOUCHIROUD, G. SABEUR & G. BERNARDI. 1993. The compositional patterns of the avian genomes and their evolutionary implications. J. Mol. Evol. **37:** 544–551.
10. MOUCHIROUD, D. & G. BERNARDI. 1993. Compositional properties of coding sequences and mammalian phylogeny. J. Mol. Evol. **37:** 109–116.
11. BERNARDI, G., B. OLOFSSON, J. FILIPSKI, M. ZERIAL, J. SALINAS, G. CUNY, M. MEUNIER-ROTIVAL & F. RODIER. 1985. The mosaic genome of warm-blooded vertebrates. Science **228:** 953–958.
12. CLAY, O., S. CACCIO, S. ZOUBAK, D. MOUCHIROUD, & G. BERNARDI. 1996. Human coding and noncoding DNA: Compositional correlations. Mol. Phylogen. Evol. **5:** 2–12.
13. BERNARDI, G. 1993. The human genome organization and its evolutionary history: A review. Gene **135:** 57–66.
14. MOUCHIROUD, D., G. D'ONOFRIO, B. AÏSSANI, G. MACAYA, C. GAUTIER & G. BERNARDI. 1991. The distribution of genes in the human genome. Gene **100:** 181–187.
15. AÏSSANI, B. & G. BERNARDI. 1991. CpG islands, genes, isochores in the genome of vertebrates. Gene **106:** 185–195.
16. AÏSSANI, B. & G. BERNARDI. 1991. CpG islands features and distribution in the genome of vertebrates. Gene **106:** 173–183.
17. JABBARI, K. & G. BERNARDI. 1998. CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families. Gene **224:** 123–127.
18. KEREM, B.S., R. GOITEN, G. DIAMOND, H. CEDAR & M. MARCUS. 1984. Mapping of DNAase I sensitive regions of mitotic chromosomes. Cell **38:** 493–499.
19. TAZI, J. & A. BIRD. 1990. Alternative chromatin structure at CpG islands. Cell **60:** 909–920.
20. SACCONE, S., A. DE SARIO, G. DELLA VALLE & G. BERNARDI. 1992. The highest gene concentrations in the human genome are in T bands of metaphase chromosomes. Proc. Natl. Acad. Sci. USA **89:** 4913–4917.
21. SACCONE, S., A. DE SARIO, J. WIEGANT, A.K. RAAP, G. DELLA VALLE & G. BERNARDI. 1993. Correlations between isochores and chromosomal bands in the human genome. Proc. Natl. Acad. Sci. USA **90:** 11929–11933.
22. SACCONE, S., C. CACCIO, J. KUSUDA, L. ANDREOZZI & G. BERNARDI. 1996. Identification of the gene-richest bands in human chromosomes. Gene **174:** 85–94.
23. FEDERICO, C., S. SACCONE & G. BERNARDI. 1998. The gene-richest bands of human chromosomes replicate at the onset of the S-phase. Cytogenet & Cell Genet. **80:** 83–88.
24. BERNARDI, G., S. HUGHES & D. MOUCHIROUD. 1997. The major compositional transitions in the vertebrate genome. J. Mol. Evol. **44:** S44–S51.
25. BERNARDI, G. & G. BERNARDI. 1990. Compositional patterns in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. **31:** 265–281.
26. BERNARDI, G. & G. BERNARDI. 1990. Compositional transitions in the nuclear genomes of cold-blooded vertebrates. J. Mol. Evol. **31:** 282–293.
27. BERNARDI, G. & G. BERNARDI. 1991. Compositional properties of nuclear genes from cold-blooded vertebrates. J. Mol. Evol. **33:** 57–67.
28. DUTRILLAUX, B. 1973. Nouveau système de marquage chromosomique: Les bandes T. Chromosoma **41:** 395–402.
29. HASEGAWA, S., T. YASUNAGA & T. MIYATA. 1979. Secondary structure of MS2 phage RNA and bias in code word usage. Nucleic Acids Res. **7:** 2073–2079.
30. WADA, A. & A. SUYAMA. 1986. Local stability of DNA and RNA secondary structure and its relation to biological functions. Prog. Biophys. Molec. Biol. **47:** 113–157.
31. GALTIER, N. & J.R. LOBRY. 1997. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. **44:** 632–636.

32. D'ONOFRIO, G., D. MOUCHIROUD, B. AÏSSANI, C. GAUTIER & G. BERNARDI. 1991. Correlations between the compositional properties of human genes, codon usage and amino acid composition of proteins. J. Mol. Evol. **32:** 504–510.
33. D'ONOFRIO, G. & G. BERNARDI. 1992. A universal compositional correlation among codon positions. Gene **110:** 81–88.
34. CACCIO, S., S. ZOUBAK, G. D'ONOFRIO & G. BERNARDI. 1995. Nonrandom frequency patterns of synonymous substitutions in homologous mammalian genes. J. Mol. Evol. **40**: 280–292.
35. MOUCHIROUD, D., C. GAUTIER & G. BERNARDI. 1995. Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of non-synonymous substitutions. J. Mol. Evol. **40:** 107–113.
36. ZOUBAK, S., G. D'ONOFRIO, S. CACCIO, G. BERNARDI & G. BERNARDI. 1995. Specific compositional patterns of synonymous positions in homologous mammalian genes. J. Mol. Evol. **40:** 293–307.
37. ALVAREZ-VALIN, F., K. JABBARI & G. BERNARDI. 1998. Synonymous and nonsynonymous substitutions in mammalian genes: Intragenic correlation. J. Mol. Evol. **46:** 37–44.
38. AKASHI, H. 1994. Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translation accuracy. Genetics **136:** 927–935.
39. LI, W.-H. 1997. Molecular Evolution. Sinauer Associated, Inc., Publishers. Sunderland, MA.
40. ROBINSON H., Y.G. GAO, B.S. MCCRARY, S.P. Edmondson, J.W. Shriver & A.H. Wang. 1998. The hyperthermophyle chromosomal protein Sac7d sharply kinks DNA. Nature **392:** 202–205.
41. TAGUCHI, H., J. KONISHI, N. ISHII & M. YOSHIDA. 1991. A chaperonin from a thermophilic bacterium *Thermus termophylus*, that controls refolding of several thermophilic enzymes. J. Biol. Chem. **266:** 22411–22418.
42. DURET, L., D. MOUCHIROUD & M. GOUY. 1995. Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J. Mol. Evol. **40:** 308–317.
43. GRANTHAM, R. 1980. Workings of the genetic code. Trends Biochem. Sci. **5:** 327–333.
44. KYTE, J. & R.F. DOOLITTLE. 1982. A simple method for displaiyng hydropathic character of a protein. J. Mol. Biol. **157:** 105–132.