J Mol Evol (1987) 26:198–204

Journal of Molecular Evolution

Compositional Compartmentalization and Gene Composition in the Genome of Vertebrates

Dominique Mouchiroud,[1] Gwennaele Fichant,[1] and Giorgio Bernardi[2]

[1] Laboratoire de Biométrie, U.A. 243, Université Claude Bernard Lyon I, 69622 Villeurbanne, France
[2] Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 75005 Paris, France

Summary. The compositional distribution of coding sequences from five vertebrates (*Xenopus*, chicken, mouse, rat, and human) is shifted toward higher GC values compared to that of the DNA molecules (in the 35–85-kb size range) isolated from the corresponding genomes. This shift is due to the lower GC levels of intergenic sequences compared to coding sequences. In the cold-blooded vertebrate, the two distributions are similar in that GC-poor genes and GC-poor DNA molecules are largely predominant. In contrast, in the warm-blooded vertebrates, GC-rich genes are largely predominant over GC-poor genes, whereas GC-poor DNA molecules are largely predominant over GC-rich DNA molecules. As a consequence, the genomes of warm-blooded vertebrates show a compositional gradient of gene concentration. The compositional distributions of coding sequences (as well as of DNA molecules) showed remarkable differences between chicken and mammals, and between mouse (or rat) and human. Differences were also detected in the compositional distribution of housekeeping and tissue-specific genes, the former being more abundant among GC-rich genes.

Key words: Genome — Isochores — Gene composition — Coding sequences — Vertebrates

Introduction

A breakthrough has recently been made in our understanding of the organization and evolution of the vertebrate genome, due to some findings (Bernardi et al. 1985) that can be summarized as follows. (1) The nuclear genome of warm-blooded vertebrates is characterized by a striking compositional compartmentalization in that it consists of a mosaic of very long (>300 kb) DNA segments, called isochores, which belong to a small number of classes showing different GC levels and a fairly homogeneous base composition, at least in the 3–300-kb range. (2) GC-poor and GC-rich isochores appear to correspond to the DNA segments present in Giemsa and Reverse chromosomal bands, respectively. (3) The localization of a number of sequences in the isochore classes (or genome compartments) of warm-blooded vertebrates has revealed that the distribution of genes is strongly biased toward GC-rich isochores and tends to be conserved in evolution. (4) The GC levels of both coding and noncoding sequences (e.g., introns, specific families of interspersed repeats, etc.), as well as of codon third positions and CpG doublet levels show a linear dependence upon the GC levels of the genome compartments harboring the sequences. (5) In contrast with the situation just described for warm-blooded vertebrates, GC-rich isochores are either absent or poorly represented in the genomes of the vast majority of cold-blooded vertebrates. These genomes are much less compartmentalized, and Giemsa and Reverse chromosomal bands are poor or absent (Medrano et al. 1987). Some general issues related to the results obtained on the vertebrate genomes have been recently discussed elsewhere (Bernardi and Bernardi 1985, 1986a,b; Mouchiroud 1986).

The basic findings outlined above could only be obtained by combining DNA fractionation in density gradients run in the presence of sequence-spe-

Offprint requests to: G. Bernardi

cific ligands with the localization of genes in the fractions so obtained. This laborious approach can hardly be extended to a large number of genes, and yet such an extension is needed in order to reach a general view of genome organization in vertebrates. Alternative pathways to gather additional information exist, however. These comprise (1) investigating the GC content and other sequence properties (such as CpG levels or GC contents of codon third positions) of genes from vertebrate genomes; (2) scanning the GC levels of extended DNA sequences; and (3) correlating the GC content of the genes of vertebrates with their chromosomal band location or with their replication timing (the latter is early and late in the cell cycle for Reverse and Giemsa bands, respectively). The first approach was used to show that the compositional distribution of coding sequences was strikingly different for 19 genes from cold-blooded vertebrates, which were predominantly GC-poor, and for 34 human genes, which showed a wide GC level spectrum and a predominance of GC-rich genes (Bernardi et al. 1985). Using the second approach, it was shown that long DNA regions surrounding genes were indeed fairly homogeneous in base composition, the GC levels being close to those of the surrounding genes (Aota and Ikemura 1986). The third approach showed that, in humans, GC-poor and GC-rich genes were mostly present in Giemsa and Reverse bands, respectively (Aota and Ikemura 1986).

When considering the compositional distribution of genes from warm-blooded vertebrates, an additional element to be taken into consideration is the suggestion that housekeeping genes have a different distribution in chromosomal bands, compared to tissue-specific genes. Indeed, some of the former genes (which fulfill basic cell functions and are, therefore, expressed in every cell) were localized in Reverse bands, whereas some of the latter (which are expressed in a tissue-specific and/or in a developmentally regulated way) were localized in Giemsa bands (Goldman et al. 1984). This led to the proposal of the existence of two genomes—a housekeeping genome, corresponding to Reverse bands, and an ontogenetic genome, corresponding to Giemsa bands (Goldman et al. 1984; Holmquist 1987). The gene localization results of Bernardi et al. (1985) indicated, however, that while the few housekeeping genes tested were restricted to GC-rich components in warm-blooded vertebrates, tissue-specific genes could be distributed in both GC-poor and GC-rich compartments.

Here we present an investigation on the GC distribution of a much larger sample of vertebrate genes compared to those studied so far, focusing on the five species on which information is most abundant: human, rat, mouse, chicken, and *Xenopus*. The main questions we addressed concern (1) the correlation between the compositional distributions of genes and DNA molecules; (2) the differences in gene distributions (a) between cold-blooded and warm-blooded vertebrates, (b) between chicken and mammals, and (c) between mouse (or rat) and man; and (3) the compositional distribution of housekeeping and tissue-specific genes.

Materials and Methods

Compositional distributions of DNA molecules from *Xenopus* (*Xenopus laevis*), chicken, mouse, and human were those reported by previous investigations on the "major DNA components," namely on the DNA fragments derived from each of the isochore classes of these genomes (Thiery et al. 1976; Cortadas et al. 1979; Cuny et al. 1981; Olofsson and Bernardi 1983; Salinas et al. 1986; Zerial et al. 1986). Sedimentation coefficients for the DNA preparations used were 35*S* (*Xenopus*), 47*S* (chicken), 39*S* (mouse), and 32*S* (human). Using the relationship of Eigner and Doty (1965), these sedimentation coefficients correspond to molecular weights of 27, 57, 36, and 22 million daltons, respectively. Fractionation procedures involved preparative Cs_2SO_4/Ag^+ (*Xenopus*, chicken, mouse, human) or Cs_2SO_4/BAMD [chicken, mouse, human; BAMD is 3,6 bis (acetato-mercuri methyl) dioxane] density gradient centrifugations (see the references quoted above for the experimental conditions). In the case of the chicken genome, the 1.712 g/cm^3 minor component, corresponding to 3% of total DNA, was considered here as a major component, which was called H3 by analogy with the H3 component of the human genome (Bernardi et al. 1985). Recent results (Sabeur, Filipski, and Bernardi, in preparation) on rat DNA indicate a great similarity with mouse DNA, as far as "main-band" DNA is concerned.

For most density components, GC contents were determined by nucleoside analysis; otherwise, they were derived from buoyant densities in CsCl, using the relationship of Schildkraut et al. (1962). The intermolecular compositional heterogeneity, H, of major DNA components was estimated according to Schmid and Hearst (1972). H is equal to the standard deviation of the compositional distribution curve.

GC contents of protein-coding genes were obtained from GenBank, Release 48 (February 1987). The ACNUC retrieval system (Gouy et al. 1984) was used. Computations were done using ANALSEQ, a software for statistical analysis of genetic sequences developed by the Laboratoire de Biométrie in Lyon. Listings of the genes investigated will be provided upon request. Coding regions were considered from the initial AUG to the termination codon. In the case of *Xenopus*, data concerning genes not yet available in GenBank were obtained from the literature. These genes were: pre-proopiomelanocortin (Martens 1987), ribosomal protein L14 (Beccari and Mazzetti 1987), and c-myc (Taylor et al. 1986) genes. The GC contents of the coding sequences of these genes were 45.1, 48.8, and 51.7%, respectively.

Since genes located in any given gene cluster have approximately the same GC levels, compositional distributions of clustered genes should preferably be plotted for the gene clusters and not for individual genes (Bernardi et al. 1985), in order to avoid a bias. This has been done here whenever gene clustering was known to be present.

Results

Figure 1 depicts a comparison of the compositional distribution of DNA molecules (in the 35–85-kb size

range) and of coding sequences from *Xenopus*, chicken, mouse, rat, and man. A window of 2.5% GC was chosen as most convenient for these comparisons, which involved gene numbers ranging from 23 to 245. Since each major DNA component has a compositional heterogeneity of 2.5–3.8% GC (Cuny et al. 1981; Olofsson and Bernardi 1983), the 2.5% GC bars only correspond to the centers of distribution of major components.

Three main features are immediately evident in the histograms of Fig. 1. (1) In all cases, the gene distribution shows an overall shift toward higher GC levels compared to the distribution of DNA molecules. (2) The distribution of the *Xenopus* genes tested is similar to the distribution of the DNA molecules, in that (a) both DNA and gene distributions are narrow; and (b) GC-poor genes are largely predominant, as are GC-poor DNA molecules. (3) In the case of warm-blooded vertebrates, the gene distribution is wide and multimodal; moreover, GC-rich genes are more abundant than GC-poor genes, whereas GC-rich DNA molecules are less abundant than GC-poor DNA molecules; in other words, the two distributions are quite different and are almost mirror images of each other.

The distribution of the *Xenopus* genes only comprised 23 genes that were centered around the GC level of the GC-poor peak(s) of the mammalian genes, with the exceptions of the histone H4 and the cytokeratin genes.

In the chicken genome, the distribution involved 56 genes and appeared to be multimodal with a predominant GC-rich peak. The lower limit of the distribution was coincident with those of mammals, but the upper limit was definitely higher than those of rodents and also of humans if account is taken of the different numbers of coding sequences.

In the case of the human genome, the compositional distribution of coding sequences concerned 245 genes covering a 37.5–75% GC range and was multimodal with two major peaks at about 46% and 58% GC, the second peak being predominant. In the case of the mouse genome, the distribution included 134 genes. While the lower limit of the distribution was identical with that of the human genes, the upper limit was lower. Moreover, the overall compositional distribution of genes was different from that of human genes in that the predominance of GC-rich genes was less marked. The distribution of the 144 coding sequences from the rat genome was similar to, but narrower than, that of mouse.

The compositional distribution of genes from the three genomes which comprise the largest number of sequenced genes, those of human, mouse, and rat, were also studied by using a narrower, 1.5% GC window, to obtain a higher resolution. This revealed the existence of a fine structure in the distribution

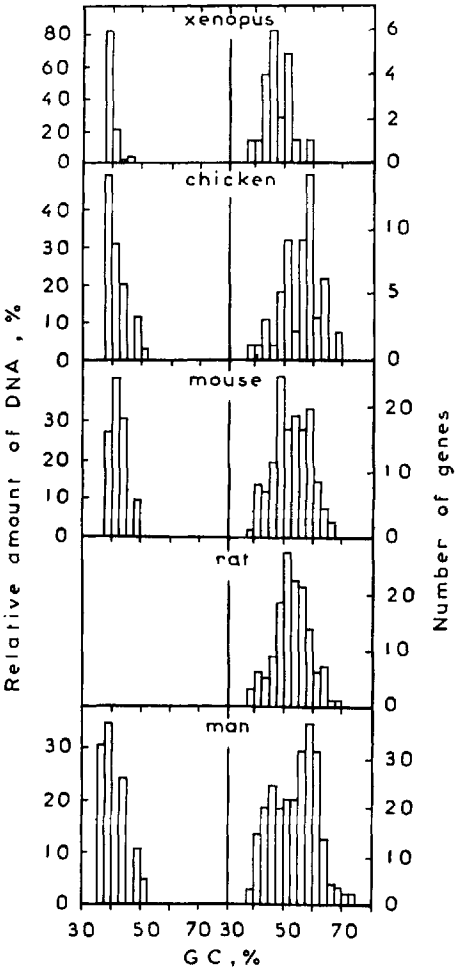

Fig. 1. Relative amounts of major DNA components (lefthand frames; in the case of warm-blooded vertebrates bars correspond to major components L1, L2, H1, H2, and, if present, H3; data from Bernardi et al. 1985; the DNA distribution of rat is similar to that of mouse; Sabeur, Filipski, and Bernardi, in preparation) and number of genes (righthand frames) from vertebrate genomes are plotted against the GC levels of major DNA components and of coding regions, respectively. A 2.5% GC window was used; since major components from vertebrate DNAs range in H values (see Materials and Methods) from 2.5 to 3.8% GC, bars on the lefthand frames only correspond to the centers of their compositional distributions. Numbers of genes explored were 23, 56, 134, 144, and 245 for *Xenopus*, chicken, mouse, rat, and man, respectively.

of coding sequences from both genomes and confirmed the existence of differences between the distributions of rodent and human genes (Fig. 2).

The main features exhibited by the compositional distributions of the coding sequences (Figs. 1 and 2) were confirmed and made more evident by comparing the compositional distributions of codon third positions (Fig. 3).

Finally, housekeeping genes from warm-blooded vertebrates were found among both GC-poor and GC-rich genes (Fig. 4), but a higher abundance among GC-rich genes could be seen (see also Table 1). This situation was also evident in *Xenopus*.

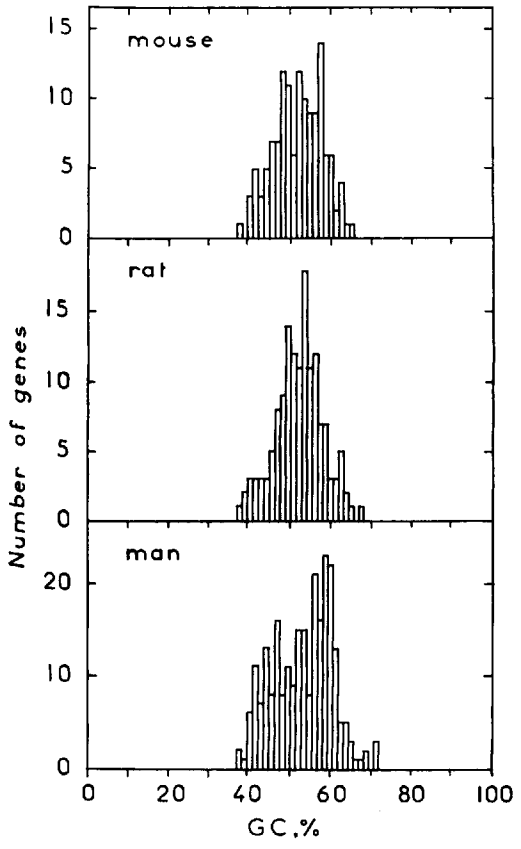

Fig. 2. The numbers of mouse, rat, and human genes are plotted against the GC levels of the coding regions. A 1.5% GC window was used.

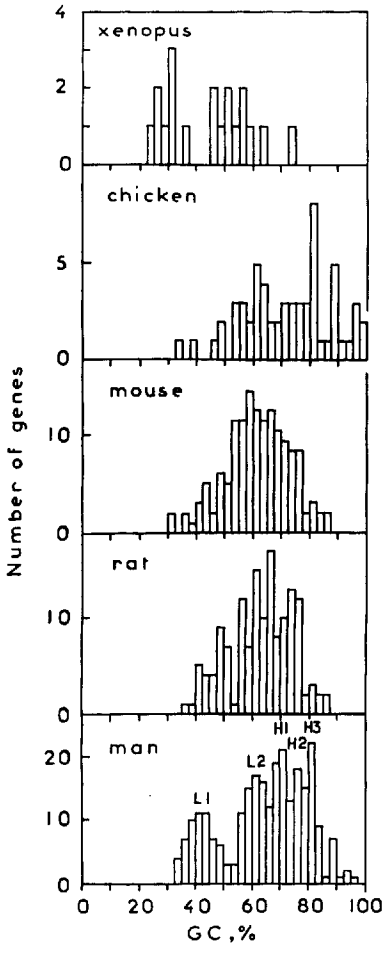

Fig. 3. The numbers of genes are plotted against the GC levels of codon third positions. A 2.5% GC window was used. Tentative identifications of the compartments of the human genome (L1, L2, H1, H2, H3) are indicated (see text).

Discussion

The Compositional Distribution of Vertebrate Genes (Coding Sequences)

The main conclusions reached are the following:

1) Coding sequences from vertebrate genomes exhibit a compositional distribution which is shifted toward higher GC levels compared to that of the corresponding DNA molecules. This result is explained by the fact that intergenic noncoding sequences (which form the major part of the vertebrate genomes) are lower in GC compared to coding sequences (Bernardi et al. 1985; Bernardi and Bernardi 1986b; and manuscript in preparation). It should be noted that coding sequences were chosen because of their large numbers in GenBank. Very similar results would, however, be obtained for exon and intron sequences, since introns show only a slightly lower GC level compared to exons (Bernardi et al. 1985; Bernardi and Bernardi 1986a,b; and manuscript in preparation).

2) The compositional distribution of the 23 genes from the *Xenopus* genome is similar to that of DNA molecules. Indeed, most coding sequences and most DNA molecules are GC-poor. This result is based on an admittedly small gene sample, but it can, however, be trusted (a) on the basis of the fact that *all* available *Xenopus* genes, and not a selected sample, were used (this was equally the case for all other genes examined here); (b) because it is supported by data on genes from other cold-blooded vertebrates (Bernardi et al. 1985; Perrin and Bernardi 1987); and (c) because of the lack or very low level of GC-rich DNA components in the genomes of cold-blooded vertebrates. It should be noted that *Xenopus* is one of the rare cold-blooded vertebrates (see Hudson et al. 1980; Medrano et al. 1987) that has a relatively heterogeneous genome. The majority of cold-blooded vertebrates, which have a more homogeneous genome, can, therefore, be expected to show gene and DNA distributions even more similar to each other.

3) In contrast to the previous case, the compositional distributions of DNA molecules and coding sequences from warm-blooded vertebrates are different. The former is characterized by a predominance of GC-poor molecules, the latter by that of GC-rich genes. The difference, seen in the histo-

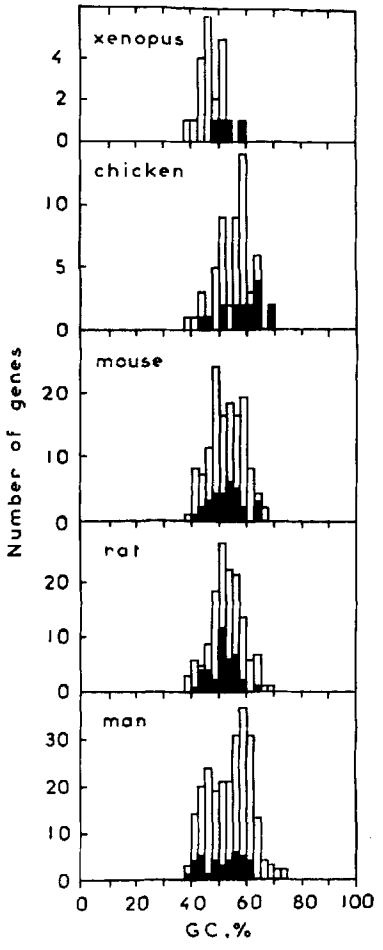

Fig. 4. The numbers of housekeeping genes (black bars) and of all genes investigated (white bars) are plotted against GC levels of coding regions. A 2.5% window was used

grams of Fig. 1, may be even more pronounced in reality, because of the current underrepresentation in GenBank of housekeeping genes, which tend to be more abundant among GC-rich genes (see below).

It is interesting to note that the higher GC level of coding sequences compared to intergenic sequences (and, to a lesser extent, to introns) and the predominance of GC-rich genes are the two factors that account for the early observations that most mammalian coding sequences are higher in GC than the genomes from which they are derived.

4) A novel result concerns the fact that the compositional distribution of the chicken genes extends further toward GC-rich values compared to those of mammals. Results obtained by fractionating chicken and mammalian DNAs (Thiery et al. 1976; Cortadas et al. 1979; Kadi, Filipski, and Bernardi, in preparation) show similar differences. Even if no gene has been localized so far in the GC-richest component, H3, of the chicken genome, it is most likely that some are located there. Indeed, the GC-richest coding sequences of chicken are up to 10%

Table 1. GC levels of coding sequences and codon third positions from genes localized in isochore classes

Species/gene locus[a]	% GC		Isochore class[b]
	Coding sequence	Codon third position	
Xenopus			
1 β-globin	48.8	50.4	L1–L2
α-globin	45.2	51.1	L1–L2
Chicken			
1 Ovalbumin	44.9	46.7	L1–L2
Y	42.9	48.9	L1–L2
2 Vitellogenin	47.9	58.0	L1–L2
3 αD-globin	57.5	77.4	H2
αA-globin	59.4	81.1	H2
4 β-globin	59.2	81.1	H2
ρ-globin	59.4	87.8	H2
Mouse			
1 Igk variable	46.2	43.5	L1
2 Igk constant	48.8	58.5	L2
3 α_c-actin	51.3	61.2	L2*
4 β-minor globin	55.8	66.9	L2
β-major globin	55.3	66.9	L2
5 α-globin	58.0	67.9	H2
6 α_s-actin	57.6	77.2	H2*
Man			
1 β-globin[c]	56.3	66.2	L2
2 c-mos	61.1	72.9	L2–H1*
3 c-myc	58.9	76.6	H1–H2
4 c-sis	62.3	78.0	H3*
5 c-Ha-ras 1	59.3	80.2	H3*
6 α1-globin	65.0	88.8	H3*
α2-globin	65.0	88.8	H3*
7 p-omc	67.8	90.0	H3*

References for gene sequences are given in Bernardi et al. (1985)

[a] Locus indicates here a gene or a gene cluster (see Bernardi et al. 1985)

[b] Slightly modified from Bernardi et al. (1985), Salinas et al. (1986), and Zerial et al. (1986). Asterisked values refer to localizations in Cs_2SO_4/BAMD density gradients

[c] Average of the localized genes from the β-globin cluster

higher in GC than the α- and β-globin genes, which have been localized in the H2 component.

5) Another novel result concerns differences in the compositional distribution of mouse and human genes. These differences are in keeping with the different DNA distributions found in the genomes of mouse (Salinas et al. 1986) and man (Zerial et al. 1986). In fact, the mouse genome does not contain the heaviest DNA component, H3, which is present in the human genome. Moreover, the predominance of the GC-rich genes is much less striking in the mouse genome than in the human genome. These features of the mouse genome are found also in the genome of another murid rodent, the rat.

The presence in human of the GC-richest H3 component (which is absent in mouse) was suggested to be either associated with a general GC

enrichment also involving genes, or to the "invasion" of "amplified" GC-rich repetitive sequences, like Alu sequences, which are five times more abundant in man compared to mouse and are particularly abundant in H3 (Salinas et al. 1986; Zerial et al. 1986). The higher relative amount of GC-rich coding sequences in man, compared to either mouse or rat, indicates that, while Alu and other GC-rich repetitive sequences may have contributed to the GC enrichment of H3, a more fundamental process involving point mutations in genes also took place (Perrin and Bernardi, in preparation). At this point, it should be mentioned that the suspected presence of H3 in the genomes of mammals from orders other than primates and rodents has been confirmed for the five orders explored so far (Sabeur, Filipski, and Bernardi, in preparation). If this situation also holds for the mammalian orders not yet tested (in particular for insectivores and chiropters), the case for an ancient separate evolutionary origin of murid rodents would receive support at the molecular level. The remarkable differences in the compositional distribution of genes between murid rodents and man found here would be in keeping with this idea.

6) When a higher resolution was used to investigate the compositional distribution of coding sequences from mouse, rat, and human genes (Fig. 2), a fine structure was detected that confirmed the differences between the compositional distributions of human and rodent genes as seen at a lower resolution. In man, the major peak is higher in GC, and both GC-poor and GC-rich peaks are more numerous and better individualized; moreover, GC-rich peaks reach higher GC values. The more complex compositional distribution of human genes compared to rodent genes appears to be highly significant in view of the large gene samples analyzed. As far as the differences between mouse and rat gene distribution are concerned, it should be mentioned that homologous coding sequences only represented a small minority of mouse and rat genes. Differences are therefore likely to be largely due to the different gene samples used in the two cases.

7) All the conclusions drawn above on the basis of the compositional distribution of coding sequences are confirmed by those of codon third positions. Because of the elimination of the less variable GC levels of second and first positions, the histograms span a wider range (in the case of the human genes from 32.5 to 97.5% GC) and results are more clearcut (Fig. 3). A further improvement could be expected by using the third positions of quartet codons, but the advantages of these plots are counterbalanced by the reduction (by about 50%) in the number of codons analyzed.

Finally, the results concerning the compositional distributions of genes point to the fact that these distributions represent a new extremely useful tool in phylogenetic studies; its importance will obviously increase with the increase in the number of coding sequences available in data banks.

The Compositional Distribution of Vertebrate Genes and the Genome Compartments

The linear relationships found between the GC levels of coding sequences and codon third positions and those of the genome compartments harboring them suggest that the compositional distribution of coding sequences and codon third positions should reflect that of genome compartments. The GC levels of localized coding sequences (Table 1) are in basic agreement with this expectation, allowance being made for a certain overlap of distribution curves. Indeed, GC levels of coding sequences and more so of codon third positions are higher for genes localized in compartments of higher GC levels. If one takes into account the GC levels of codon third positions from genes localized in isochore classes (Table 1), one can crudely set the limit between the coding sequences located in the two light (L1–L2) and those in the two heavy compartments (H1–H2) at 67% GC, and the limit betweeen the coding sequences located in the latter and those comprised in the heaviest H3 compartment at 77% GC (Fig. 1). Indeed, on this basis, one might even tentatively identify the major DNA components of the human genome with the peaks shown in Fig. 3; more localization results of known coding sequences are, however, needed to make this identification certain. Incidentally, the strong heterogeneity of the H3 component indicated by the histogram of Fig. 3 is also suggested by investigations at the DNA level (Zerial et al. 1986).

The limits between genome compartments can, in turn, be used to approximately estimate the relative numbers of genes in the compartments of the human genome. In this case, the codon third position data suggest that approximately 48% of genes are located in the L1–L2 compartments, 32% in the H1–H2 compartments, and 20% in the H3 compartment. Even if, admittedly, these estimates are rather crude, they are good enough to confirm, on a much larger gene sample, the striking nonuniformity of gene distribution in the genome of warm-blooded vertebrates reported by Bernardi et al. (1985). Indeed, the three genome sections under consideration correspond to about 65%, 32%, and 3% of the human genome. The data mentioned above clearly point to the existence of a very steep compositional gradient of gene concentration in the human genome, genes being roughly 10 times more frequent in the heaviest, H3, component than in the

two lightest ones, L1–L2. In view of the current underrepresentation in GenBank of housekeeping genes (which correspond to the vast majority of genes and are predominantly GC-rich), the gene gradient is likely to be much steeper in reality.

As far as the compositional gradient of genes from warm-blooded vertebrates is concerned, a possible explanation (Bernardi and Bernardi 1986b) is that the regional enrichment in GC, which led to the formation of GC-rich isochores at the time of the transition between cold-blooded and warm-blooded vertebrates, initially took place in and around genes (and more particularly around housekeeping genes). Needless to say, the existence of a gene gradient in warm-blooded vertebrates also has a practical importance insofar as the human genome project (Palca 1987) is concerned, since mapping and sequencing will be much more rewarding in the H3 compartment than in the other genome compartments.

The Compositional Distribution of Housekeeping and Tissue-Specific Genes

The results in Fig. 4 suggest that housekeeping genes are more abundant among GC-rich than among GC-poor genes. Housekeeping genes definitely are well represented, however, among GC-poor genes. The presence of housekeeping genes among GC-poor genes (as well as the previously demonstrated presence of tissue-specific genes among GC-rich genes; Bernardi et al. 1985) contradicts the extreme view of an exclusive localization of housekeeping genes in Reverse bands and of tissue-specific genes in Giemsa bands (Goldman et al. 1984; Holmquist 1987). The original suggestion of two genomes, a housekeeping one and an ontogenetic one, needs, therefore, to be revised in that this sharp distinction is not warranted.

Acknowledgments. The authors thank Prof. R. Grantham for his interest and encouragement, as well as for suggestions and critical reading, and Giacomo Bernardi for discussions and help in the preparation of this manuscript.

References

Aota S-I, Ikemura T (1986) Diversity in G+C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res 14:6345–6355

Beccari E, Mazzetti P (1987) The nucleotide sequence of the ribosomal protein L14 gene of *Xenopus laevis*. Nucleic Acids Res 15:1870–1872

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

Bernardi G, Bernardi G (1986a) The human genome and its evolutionary context. Cold Spring Harbor Symp Quant Biol 51:479–487

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

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

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

Cuny G, Soriano P, Macaya G, Bernardi G (1981) The major components of the mouse and human genomes. I. Preparation, basic properties and compositional heterogeneity. Eur J Biochem 115:227–233

Eigner J, Doty P (1965) The native, denatured and renatured states of deoxyribonucleic acid. J Mol Biol 12:549–580

Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A (1984) Replication timing of genes and middle repetitive sequences. Science 224:686–692

Gouy M, Milleret F, Mugnier C, Jacobzone M, Gautier C (1984) ACNUC: a nucleic acid sequence data base and analysis system. Nucleic Acids Res 12:121–127

Holmquist GP (1987) DNA sequences in G-bands and R-bands. In: Adolph KW (ed) Chromosomes and chromatin structure. CRC Press, Boca Raton FL (in press)

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

Martens GJM (1987) Structural organization of the proopiomelanocortin gene in *Xenopus laevis*: 5′ end homologies within the toad and mammalian genes. Eur J Biochem 165:467–472

Medrano L, Couturier J, Dutrillaux B, Bernardi G, Bernardi G (1987) Chromosome banding and genome compartmentalization in fishes. Chromosoma (in press)

Mouchiroud D (1986) Relation entre la composition en base de l'ADN non codant du gène et la composition en codon. C R Acad Sci (Paris) 303:743–748

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

Palca J (1987) Human genome sequencing plan wins unanimous approval in US. Nature 326:429

Pezzin P, Bernardi G (1987) Directional fixation of mutations in vertebrate evolution. J Mol Ev (in press)

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

Schildkraut CL, Marmur J, Doty P (1962) Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J Mol Biol 4:430–443

Schmid CW, Hearst JE (1972) Sedimentation equilibrium of DNA samples heterogeneous in density. Biopolymers 1:1913–1918

Taylor MV, Gusse M, Evan GI, Dathan N, Mechali M (1986) *Xenopus myc* proto-oncogene during development: expression as a stable maternal mRNA uncoupled from cell division. EMBO J 5:3563–3570

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

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

Received May 2, 1987/Revised and accepted August 4, 1987