

THE ISOCHORE ORGANIZATION OF THE HUMAN GENOME

Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005
Paris, France

CONTENTS

INTRODUCTION.....	637
ISOCHOIRES AND GENOME ORGANIZATION.....	638
<i>The Isochoires</i>	638
<i>Isochoires and Genes</i>	641
<i>Isochoires, CpG Doublets and "CpG Islands"</i>	645
<i>Isochoires, "Single-Copy" and Middle-Repetitive Sequences</i>	645
<i>Isochoires and Chromosome Bands of Metaphase Chromosomes</i>	646
<i>Isochoires and the Fine Structure of Chromosomes</i>	647
<i>Isochoires and Chromomeres</i>	648
<i>Isochoires, Chromatin Loops and Nucleosomes</i>	648
<i>Isochoires and Long-Range Physical Mapping</i>	649
ISOCHOIRES AND GENOME FUNCTIONS.....	649
<i>Isochoires and Integration of Mobile and Viral DNA Sequences</i>	649
<i>Isochoires, Translocation Breakpoints and Fragile Sites</i>	650
<i>Isochoires and Replication/Condensation Timing in the Cell Cycle</i> ..	650
<i>Isochoires, Genomic Distribution of Genes and Codon Usage</i>	651
<i>Isochoires, CpG Doublets and "CpG Islands"</i>	652
ISOCHOIRES AND GENOME EVOLUTION.....	652
CONCLUSIONS: THE PALEOGENOME AND THE NEOGENOME.....	653
PERSPECTIVES: COMPOSITIONAL MAPPING.....	655

INTRODUCTION

At a time when the human genome begins to be investigated on a massive scale, at least in the United States (19), it may be appropriate to review briefly what is already known about its organization. This information is relevant for

defining priorities and formulating strategies concerning future work, a subject that still is under debate (54, 107).

However fast our progress over the past few years, it is not generally appreciated that only about 600 human coding sequences are presently known in their primary structure. These represents 1% or 2% of all coding sequences if the total number of genes is 60,000 or 30,000, respectively, and if their average size is 1,000 base pairs and only about 0.02% of the human haploid genome, which comprises about 3 billion base pairs. Even less is known on the primary structure of noncoding, particularly intergenic, sequences. These account for 98–99% of the human genome (if the number and the size of coding sequences are those indicated above), and contain large families of repeated interspersed sequences, LINES and SINES (93), that are present in about 100,000 and 900,000 copies, respectively (44).

Consequently, it is not surprising that the recent advances in human molecular genetics have not shed much light on the general issue of genome organization, nor that many biologists still visualize the human genome, and the eukaryotic genome in general, as a “bean bag” (62), a collection of genes randomly scattered over vast expanses of “junk DNA” (74). In contrast, the investigations reviewed here indicate that the eukaryotic genome is an integrated structural, functional, and evolutionary system. This view arose from a comparative study of vertebrate genomes, centered on the analysis of their compositional patterns, namely of the compositional distributions of large DNA fragments, coding sequences, and introns (see Figures 1, 2, 5, for examples).

ISOCHORES AND GENOME ORGANIZATION

The Isochores

Equilibrium centrifugation in analytical CsCl density gradient shows that DNA preparations from warm-blooded vertebrates are characterized by a strong intermolecular compositional heterogeneity, whereas those from cold-blooded vertebrates exhibit a weak heterogeneity; moreover, the former reach GC levels that are not attained by the latter (101; see Figure 1).

To investigate the compositional distribution of DNA fragments from vertebrates in more detail, we developed an experimental approach (12, 22, 24, 29, 58, 59, 76, 82, 101, 115) derived from methods originally used for isolating satellite DNAs (15, 20, 21, 57, 65). This approach consists of fractionating DNA fragments by equilibrium centrifugation in preparative Cs₂SO₄ density gradients in the presence of sequence-specific DNA ligands, like Ag⁺ or BAMD, 3,6-bis (acetato-mercuri-methyl) dioxane; (netropsin was also used by others; 40). This approach, outlined in the legend of Figure 2, allowed the identification of a small number of families of DNA fragments

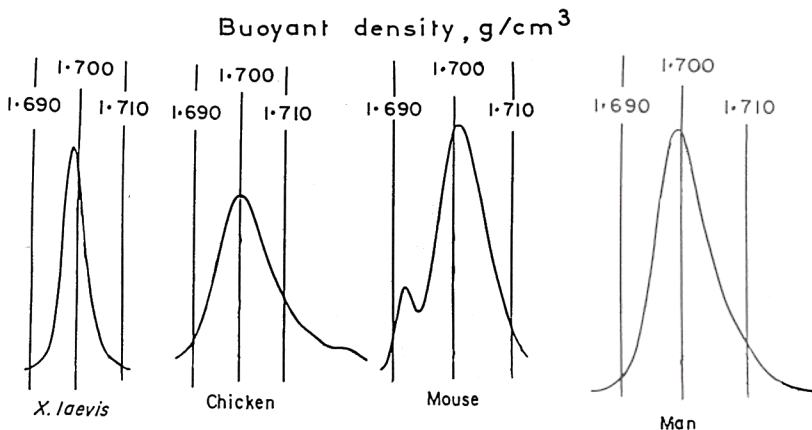


Figure 1 CsCl buoyant density profiles of DNA preparations from *Xenopus*, chicken, mouse and man. Band widths depend upon the intermolecular compositional heterogeneity (namely the spread of GC levels of DNA fragments) and molecular weight (lower molecular weights causing larger band widths because of the associated higher Brownian diffusion). After correction for differences in molecular weights (sedimentation coefficients of *Xenopus*, chicken, mouse and human DNAs were 35, 20, 24.5 and 25 S, respectively), band widths indicated a much lower intermolecular compositional heterogeneity (H; see Ref. 85) for *Xenopus* DNA (2.9% GC) than for the DNAs of chicken (5.1% GC) mouse (4.0% GC for main band) and man (4.8% GC). (Modified from ref. 101; H values are from G. Bernardi & G. Bernardi, paper in preparation) and refs. 24, 76.)

characterized by similar, though not identical, base compositions. These compositional families (or components) of DNA fragments comprise (a) major families, derived from "main band" DNA; (b) satellite families, derived from highly repeated, simple sequence satellite DNA(s); and (c) minor families, such as ribosomal DNA.

If this approach is applied to human DNA preparations in the 30–100 Kilobase (Kb) size range, one can see that the compositional distribution of DNA fragments (a) is characterized by a very broad GC range, and by the presence of two GC-poor major components, L1 and L2, representing about two thirds of the genome, and three GC-rich components, H1, H2, and H3, representing the remaining third (Figure 2; see refs. 12, 24, 101, 115); (b) is shared by the DNAs of all warm-blooded vertebrates studied, whereas the DNAs from cold-blooded vertebrates show much narrower distributions that do not extend as far in the high GC range (Figure 2); and (c) is largely independent of molecular size, at least between 3 Kb and over 300 Kb (this upper value was determined by lysing lymphocytes with sarkosyl directly in the analytical ultracentrifuge cell; the resulting very high molecular weight

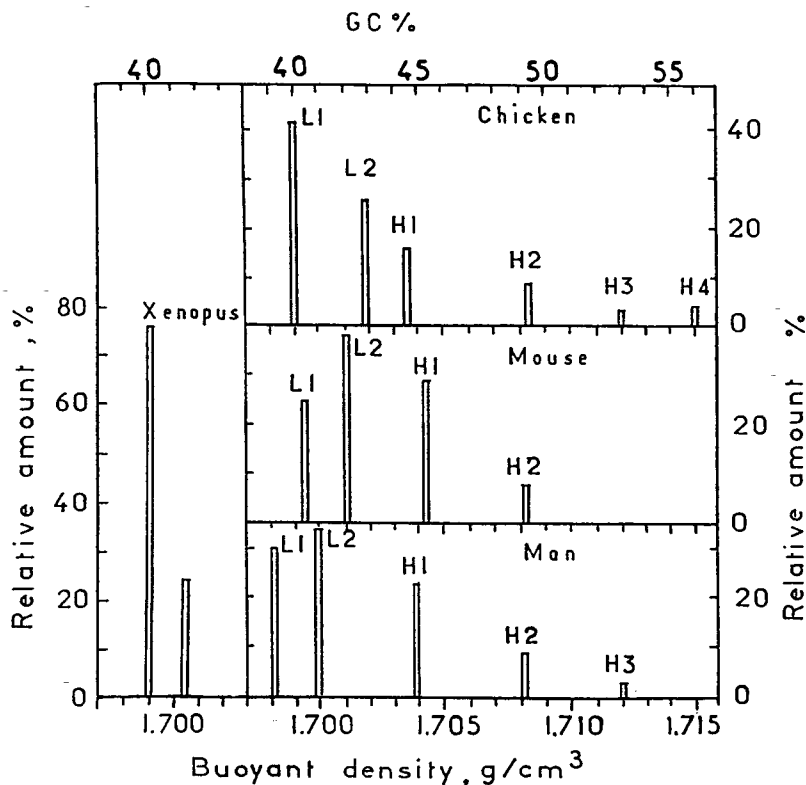


Figure 2 Histograms showing the relative amounts, modal buoyant densities and GC levels of the major DNA components from *Xenopus*, chicken, mouse and man. *Xenopus* data are from refs. 58, 101 and from G. Bernardi & G. Bernardi (paper in preparation). Chicken data (from ref. 22) include two components previously considered as minor (H3 and H4). Mouse and human data are from refs. 58, 82 and 115. Satellite and minor components are not shown. (Modified from ref. 12). The construction of the histograms involved the following steps: (a) the Cs_2SO_4 density gradient separation of DNA-ligand complexes on the basis of the sequence-specific, differential binding of the ligand (Ag^+ or BAMD) to DNA fragments of different composition; (b) the analysis of the fractions in terms of relative DNA amounts and GC levels; (c) the resolution of the analytical CsCl band profiles of DNA fractions into Gaussian curves, and the identification of a small number of families of DNA fragments that are similar in composition (see ref. 58 for further details); (d) the pooling together of fractions containing the same families, after re-centrifugation in the case of very heterogeneous fractions; and (e) the assessment of the relative DNA amounts, modal buoyant densities and GC levels of the major DNA components. It should be noted that while the CsCl profiles of the isolated major components are centered on the modal buoyant densities shown in the Figure, they exhibit some degree of overlap with each other, particularly for the components that are most abundant and closer in density.

preparations exhibited multimodal profiles with maxima corresponding to the buoyant densities of major components; 58).

The independence of the compositional distribution of DNA fragments from molecular size indicates a remarkable compositional homogeneity over very long DNA stretches, called "isochores", (for "equal regions"; 12, 24). These gave rise to the DNA fragments that were actually fractionated, because of the occurrence of unavoidable mechanical and enzymatic breakage during DNA preparation. The existence of isochores is also indicated by the very low intermolecular heterogeneities of vertebrate DNAs and their components relative to their genome size and kinetic complexities, (24, 43, 79; G. Bernardi & G. Bernardi, in preparation; bacterial DNAs were used as references) as well as by other findings (see next section and Refs. 3, 45).

Satellite and minor components may also be considered as "isochores". They are not dealt with here, however, because they represent specialized genome compartments, corresponding to centromeres and telomeres, in the case of satellite DNAs, and to nucleoli, in the case of ribosomal genes.

Isochores and Genes

A number of genes from the genome of man and other warm-blooded vertebrates were localized in major components or compositional fractions by hybridization with appropriate probes (12, 23, 25, 82, 97, 115; see Figure 3). This research revealed three points:

1. In most cases, a given gene is embedded in DNA fragments that are within 1–2% of each other in GC content. This finding provides additional, independent evidence for the compositional homogeneity of isochores, as well as for the wide-spacing of isochore borders. Indeed, since the fragments making up DNA preparations are produced by random degradation, and since the genes probed can, therefore, be located anywhere on them, the narrow compositional range of the gene-carrying fragments indicates that they are very homogeneous in base composition over sizes roughly twice as large as the fragments themselves (i.e. up to 200 Kb). Incidentally, these observations are valid not only for isolated genes, but also for clustered genes, indicating again that isochores are large in comparison to the gene clusters explored (these varied in size from 4–40 Kb). Much less frequently, the gene is found in fractions covering a wider range of GC levels; obviously, if the DNA stretches around the probed gene are compositionally heterogeneous (as when the gene is close to isochore borders), random breakage produces gene-carrying fragments of different composition.

2. Gene distribution is strikingly nonuniform in the genome, most of the genes probed having been localized in component H3, which is the most heterogeneous component and only represents 3–5% of the genome (12).

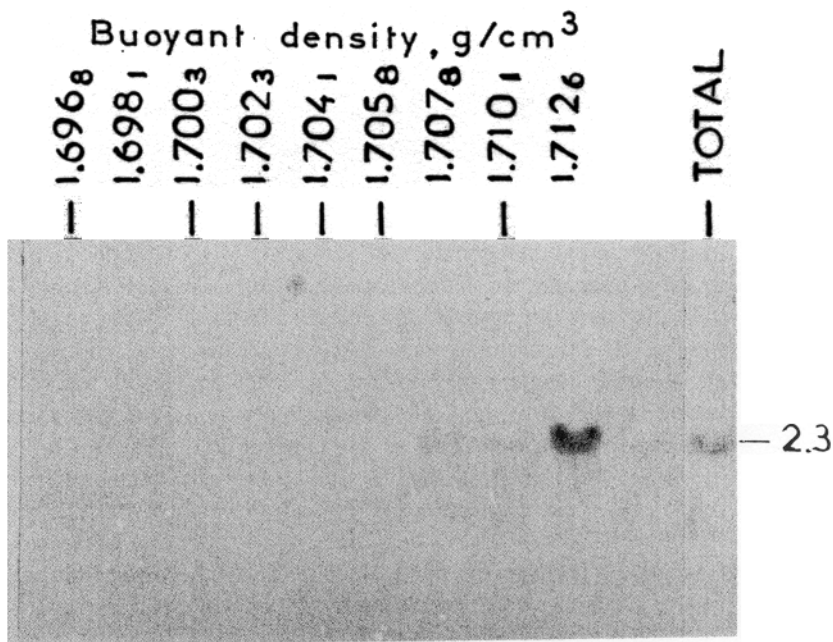


Figure 3 Localization of the *c-sis* oncogene in human DNA fractions obtained by centrifugation in a preparative Cs₂SO₄/BAMD-density gradient. After dialysis to eliminate Cs₂SO₄ and BAMD, fractions were digested with EcoRI, electrophoresed on a 0.8% agarose gel, transferred to a nitrocellulose filter and hybridized with a *c-sis* probe. A hybridization band corresponding, as expected, to a 2.3 Kb EcoRI fragment was found in a fraction having a modal buoyant density of 1.7126 g/cm³ and representing 2.3% of the human genome. (From ref. 115).

3. The GC levels of genes, introns, and exons (and their individual codon positions) are linearly correlated with those of the large DNA fragments embedding them; the GC levels of intergenic sequences are, however, systematically lower than those of exons and correspond to the unity slope passing through the origin (Figure 4). It should be noted that a correlation between GC levels of third codon positions and flanking sequences of some genes was independently reported (see 44a).

The linear relationships of Figure 4 are of interest in three respects. (a) Because genes represent only a minute amount of human DNA, isochores essentially consist of intergenic noncoding sequences. Thus the straight lines of Figure 4 correlate the GC levels of coding sequences with the GC levels of the noncoding sequences harboring them. (b) The GC levels of exons are higher by about 10% compared to those of the intergenic sequences flanking them; the slope of the exon plot is, however, equal to one. In contrast, introns, genes, and third codon position plots exhibit slopes higher than one

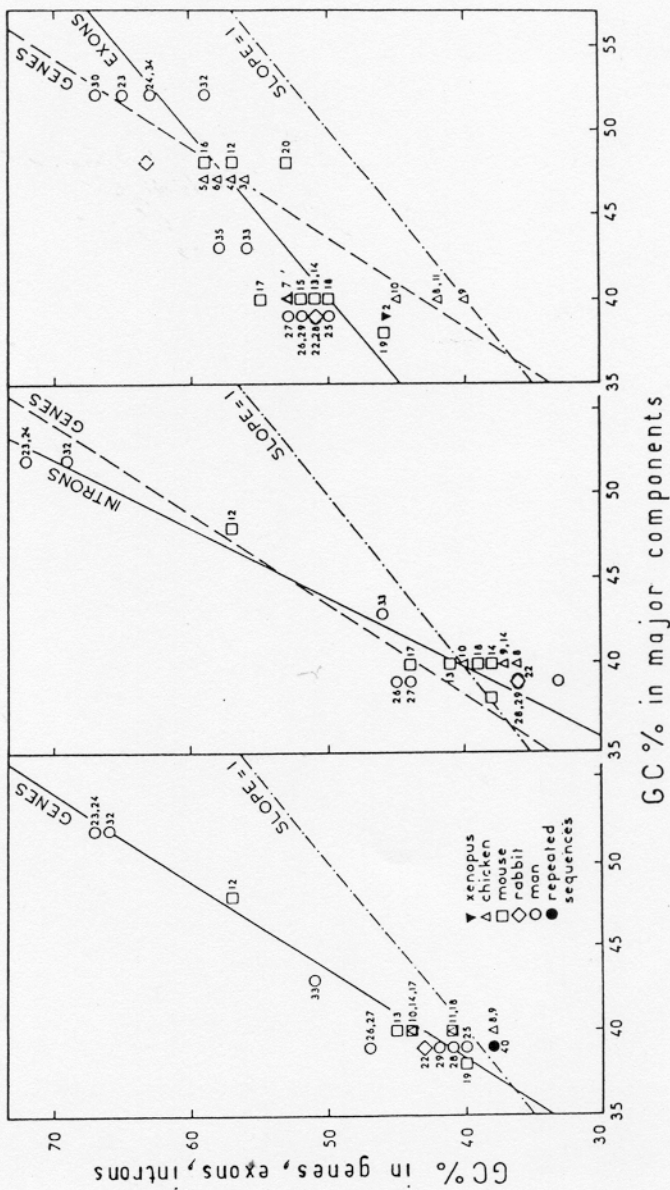


Figure 4 Plot of GC levels of genes (and mouse repeated sequence family L1), introns, and exons against the GC levels of DNA components in which they were localized. Gene localizations were made on either isolated major components or in compositional fractions (see ref. 12). In the latter case, gene localizations were assigned to the major component that was nearest in buoyant density, and this accounts for the vertical alignments of points. The numbers indicate genes and sequences (see ref. 12 for a list). Lines were drawn using the least-square method. The unit-slope lines correspond to the coincidence in GC levels of genes, introns and exons, respectively, and of the major components in which genes are located. (Modified from ref. 12).

(see Figure 4 and ref. 12). Interestingly, the different relationships just described can help in deciding whether a given anonymous probe, localized in compositional fractions, is a coding sequence or an intergenic sequence, since intergenic sequences fall on the unity slope line passing through the origin, whereas coding sequences are higher, by about 10% GC, than the DNA in which they are located. (c) The existence of such linear relationships suggests that the compositional pattern of a genome can also be studied by looking at the compositional distribution of coding sequences (and their different codon positions), and introns. The multimodal compositional patterns represented by these distributions (8-12, 70, 71, 78; see Figure 5) confirm the great predominance of GC-rich genes over GC-poor genes in the genome of man (and other warm-blooded vertebrates). In fact the concentration of genes found in the GC-richest and least abundant isochore family, H3, can be estimated as 5-10 times higher than that in the H1+H2 and L1+L2 families (70).

Isochores, CpG Doublets and "CpG Islands"

In vertebrate DNAs, CpGs are the only potential sites of methylation and the only doublets that are underrepresented relative to statistical expectations. Contrary to previous claims (80), the distribution of CpG doublets is strikingly nonuniform in the genome. Indeed, CpGs are very strongly avoided in genes and coding sequences located in GC-poor isochores, whereas they are hardly avoided at all in genes and coding sequences located in GC-rich isochores (7, 12). A similar situation is found in mammalian and avian viruses where GC-poor genomes avoid CpG, but GC-rich ones do not. Incidentally, the latter observation disposes of certain explanations for the different CpG shortage of small and large viral genomes (98, 99; see ref. 7 for further discussion).

"CpG islands" are sequences over 0.5 Kb in size, characterized by high GC levels, by clustered, unmethylated CpGs, by G/C boxes (GGGGCGGGGC and closely related sequences), and by clustered sites for rare-cutting restriction enzymes (these enzymes recognize GC-rich sequences that comprise one or two unmethylated CpG doublets; 13, 30, 33, 94, 102). CpG islands are associated with the 5' flanking sequences, exons and introns of all housekeeping genes and of many tissue-specific genes, and with the 3' exons of some tissue-specific genes (33). The distribution of CpG islands in the human and mouse genomes is also strongly nonuniform. This distribution parallels those of CpG doublets and of genes; CpG islands are rare in GC-poor isochores, but increasingly more frequent in GC-rich isochores; moreover, CpG islands are very rare in the genomes of cold-blooded vertebrates (B. Aissani, G. Bernardi & G. Bernardi, in preparation).

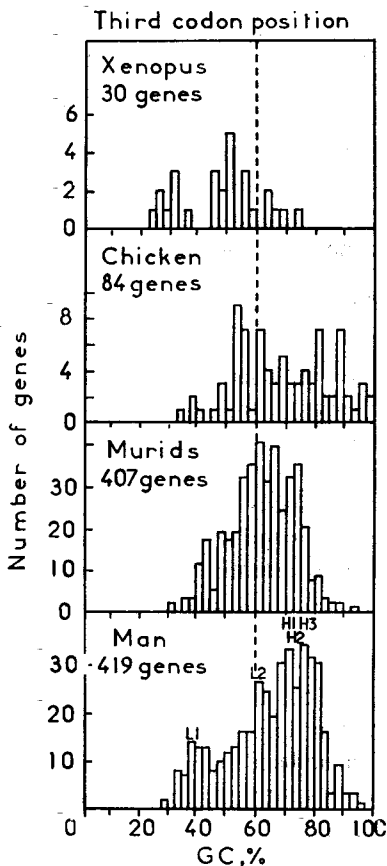


Figure 5 Compositional distribution of third codon positions from vertebrate genes. (This distribution is the most informative because of its wider spread in composition compared to coding sequences and first or second codon positions). The number of genes under consideration is indicated. The available gene sample of *Xenopus* was small, but the difference with the gene distribution from warm-blooded vertebrates was also found for homologous genes (see ref. 11), indicating that this difference is not due to the gene samples used. A 2.5% GC window was used. The broken line at 60% GC is shown to provide a reference. Approximate identifications of different compositional classes of coding sequences corresponding to the major components of the human genome (L1, L2, H1, H2 and H3) are indicated. The borders between L1-L2, H1-H2, and H3 can be tentatively estimated (70) as 67.5% and 77.5% GC, respectively. (Modified from ref. 11).

Isochores, "Single-Copy" and Middle-Repetitive Sequences

Reassociation kinetics of human major DNA components using hydroxyapatite chromatography (6, 14), revealed that "single copy" sequences decrease and middle-repetitive sequences increase from GC-poor to GC-rich components, indicating their different interspersion patterns (95); (these findings

are not inconsistent with the increase of gene concentration in GC-rich isochores since genes represent a small percentage of single-copy sequences). Moreover, different reassociating classes from the same component showed very close GC levels (95), consistent with the compositional correlations of Figure 4. These results were confirmed and corroborated by the predominant location of LINES (93) in GC-poor isochores, of SINES in the less abundant GC-rich isochores, and by the compositional match of these middle-repetitive sequences with the isochores in which they are present (66, 96).

Isochores and Chromosome Bands of Metaphase Chromosomes

G-bands (Giemsa positive or Giemsa dark bands; these are equivalent to Q-bands or Quinacrine bands), and R-bands (Reverse bands; these are equivalent to Giemsa negative, or Giemsa light bands) are produced by treating metaphase chromosomes with fluorescent dyes, proteolytic digestion, or differential denaturing conditions. GC-poor and GC-rich isochores largely correspond to the DNA of G- and R-bands, respectively. When this conclusion was proposed (24), it essentially rested on the parallelism between compositional heterogeneity of the genome, as seen at the DNA level, and chromosome banding in vertebrates. Indeed, we knew that genomes from cold-blooded vertebrates show a weak compositional heterogeneity (43, 79, 101; see Figures 1, 2, and 5), and we had realized (largely on the basis of unpublished data from other laboratories, quoted in ref. 24) that they show poor G- and R-bands, or no bands at all. In contrast, genomes from warm-blooded vertebrates exhibit a strong compositional heterogeneity (Figures 1, 2, and 5) and show distinct G- and R-bands. These findings were in agreement with the indirect cytogenetical evidence (18) suggesting that G- and R-bands correspond to GC-poor and GC-rich DNA sequences, respectively.

More recent data have confirmed this conclusion. (a) Additional evidence has been obtained for the link between compositional heterogeneity of DNA fragments and chromosome bands (63), as well as for the poor banding (1, 63, 86, 110), and the limited compositional heterogeneity of the genomes from cold-blooded vertebrates (G. Bernardi & G. Bernardi, in preparation). (b) G-bands replicate late, but R-bands replicate early (18), as do genes previously investigated in replication timing (31, 37) and localized in GC-poor and GC-rich isochores, respectively (12). (c) Genes are preferentially located in R-bands (37, 51), as well as in GC-rich isochores (11, 12, 70, 71). (d) Genes located in G- and R-bands are GC-poor and GC-rich, respectively (3, 45), as are genes located in GC-poor and GC-rich isochores (12). (e) G-bands can be produced by HaeIII degradation of metaphase chromosomes (55), because HaeIII splits GGCC sites that are much more frequent in GC-rich than in GC-poor isochores. In turn, R-bands can be produced by pancreatic DNase degradation of chromosomes protected by GC-specific binding of

chromomycin A3 (90). (f) G-bands can be obtained in metaphase chromosomes by in situ hybridization of LINES (52, 60), that are mainly located in GC-poor isochores (66, 96), and R-bands by hybridizing Alu sequences (52, 60) that are predominantly distributed in GC-rich isochores (12, 96, 115). (g) Antibodies against Z-DNA produce R-bands (104), consistent with the expectation that Z-DNA structures are more frequent in GC-rich segments. (h) Antibodies against AT-rich triplex DNA have been localized in G-bands (16). (i) Unmethylated C, present in CpG islands (predominantly located in GC-rich isochores) is mainly found in R-bands (106). (j) Gene amplification leads to the formation of homogeneous staining regions in chromosomes (87); this is the result expected if the amplified genome segments are smaller in size than isochores, as is the case.

Points made in the following sections stress the fact that the general correlation between isochores and chromosome bands should only be considered a good approximation of the actual situation. Moreover, this correlation is not true for all organisms. Indeed, an isochore organization is also present in the genomes of plants (61, 81), in which case both low and high compositional heterogeneities have been found, and yet metaphase chromosomes do not exhibit G- or R-bands. This is probably because at metaphase plant chromosomes are 3.5–6.5 times more condensed than human chromosomes (39).

Isochores and the Fine Structure of Chromosome Bands

G- and R-bands not only differ in their overall isochore make-up, but also in their internal isochore structure, as indicated in the last section and by the following points:

(a) G-bands are remarkably homogeneous in DNA composition, because they are made up of GC-poor isochores that differ very little from each other in composition (see Figure 2). In contrast, R-bands are heterogeneous, since the corresponding GC-rich isochores encompass a wide GC range. This leads to both interband and intraband heterogeneity as shown by the compositional mapping of chromosome 21 (see the final section of this review). In fact, the interspersed of different GC-rich isochores within individual R-bands was already indicated by the finding that genes located in R-bands from many chromosomes (3, 45) are present in component H3. Since the latter only represents 3–5% of the genome, it cannot account for the totality of DNA of R-bands; other components must be present. A corollary of this conclusion is that, since gene concentration is highest in the isochores of the H3 family (10, 12, 70, 71), regions of high and low gene concentration should exist not only in R- and G-bands, respectively, but also within R-bands.

(b) GC-rich and GC-poor DNA components in the human genome are in a 1:2 ratio (24, 101), whereas R- and G-bands are in a 1:1 ratio (42). This

discrepancy may mean that DNA concentration is lower in R-bands than in G-bands. In this connection, it is of interest (a) that such a lower DNA compaction, and the consequent greater accessibility of DNA, would affect chromosomal regions where the concentration of genes is higher, and (b) that DNase I sensitive chromosomal regions (corresponding to potentially active genes; 34, 109) usually correspond to Reverse bands, although not to all of them (48). An alternative explanation, not exclusive of the former, is that standard R-bands contain more GC-poor isochores (corresponding to the "thin" G-bands revealed by high resolution banding; 84, 105, 111, 112) than standard G-bands contain GC-rich isochores (corresponding to "thin" R-bands). This explanation appears to be supported by compositional mapping of chromosome 21 (see the final section) and other data (45).

ISOCHORES AND CHROMOMERES Meiotic prophase chromosomes, especially at pachytene, show a characteristic pattern of chromomeres that represent centers of chromatin condensation along the chromosomes (28, 56, 75). This pattern can be seen because meiotic chromosomes are several times more extended than mitotic chromosomes and because meiotic pairing enhances the chromomeres (18). The chromomere-interchromomere patterns of pachytene bivalents strikingly resemble the high resolution G-banding patterns visualized in mid-prophase mitotic chromosomes (84, 105, 111, 112), namely in the high-resolution banding mentioned in the preceding section. Thus, it is generally accepted that G-bands of metaphase chromosomes correspond to individual or, more frequently, to closely spaced chromomeres, and that R-bands correspond to interchromomeric regions. Consequently, it may be speculated that chromomeres and interchromomeres correspond to GC-poor and GC-rich isochores.

ISOCHORES, CHROMATIN LOOPS AND NUCLEOSOMES Levels of DNA organization lower than the chromomere-interchromomere patterns are known to exist in chromosomes. These range from the wrapping of the double helix around histones to form nucleosomes, to the packaging of nucleosomes into chromatin fibers and to the folding of the fibers into chromatin loops. The latter consist of 30–100 Kb of supercoiled DNA and are fastened at their base by non-histone scaffold proteins, mainly topoisomerase II. The scaffold-associated regions of DNA (35) may serve as preferred sites at which DNA replication begins and tend to be close to promotor elements. The possible correlations of chromatin loops with isochores are very important questions that unfortunately cannot be answered at present. Yet another open question concerns the possibility of different nucleosome spacing in GC-poor and GC-rich isochores.

Isochores and Long-Range Physical Mapping

The correlations between isochores and chromosome band structure are now being confirmed in a number of genomic regions by long-range physical mapping, made possible by the advent of pulsed field gel electrophoresis (88, 89) and of rare-cutting restriction enzymes (72).

The partial physical map (32) of the long arm of human chromosome 21 has shown that (a) genes and unique sequences are clustered in R-band q22.3 whereas they are rare in G-band q21; and (b) rare-cutting restriction enzymes cut more frequently in R-band q22.3, producing small fragments, than in G-band q21, where they yield very large fragments. Similar observations are also being made about other mapped chromosomal regions. Needless to say, these data fit with expectations based on the distribution of genes and associated "CpG islands" discussed in previous sections.

Direct evidence has been found for the expected high gene concentration in a GC-rich region, the H-2K region of the mouse major histocompatibility complex (seven genes over 150 kb) (1).

ISOCHORES AND GENOME FUNCTIONS

Isochores and Integration of Mobile and Viral DNA Sequences

Stable integration of mobile and foreign DNA sequences is mostly found in isochores of matching composition. Mobile sequences that have been amplified by retrotranscription (108) and translocated to numerous loci of the human genome during mammalian evolution, such as LINES and SINES (93), are predominantly located in isochores of matching GC levels (12, 66, 77, 95, 96, 115). This indicates that reinsertion is targeted to matching genome environments, and/or that integration is more stable within such environments. Furthermore, such reinsertion may be a source of mutations if it occurs in genes (47, 69).

Similar behavior was observed for integrated viral sequences. In the four different viral systems studied so far, bovine leukemia (BLV), hepatitis B (HBV), mouse mammary tumor (MMTV) and Rous sarcoma (RSV), integration was found predominantly in isochores of matching GC contents (49, 83, 114; F. Kadi and associates, in preparation). These isochores are very GC-rich for BLV, HBV, and RSV, and very GC-poor for MMTV. In the case of integrated MMTV sequences, endogenous sequences, which have resided in the host genome for very many generations, are more concentrated in the GC-poorest isochores than the recently integrated exogenous sequences. This finding suggests that the broader compositional distribution of recent inserts is narrowed by the selection of more stable inserts. Needless to say, these observations are of interest in connection with the integration of foreign DNA

into the genome of transgenic mammals. Here, too, an important, unresolved question concerns the effect of genomic compositional context on the expression of integrated sequences.

Isochores, Translocation Breakpoints and Fragile Sites

Translocation breakpoints are not randomly located on chromosomes (100). R-bands and G/R borders are the predominant sites of exchange processes, including spontaneous translocations, spontaneous and induced sister-chromatid exchanges, and the chromosomal abnormalities seen after X-ray and chemical damage. They also include the "hot spots" for the occurrence of mitotic chiasmata (41, 53, 68, 100,). Likewise, fragile sites tend to be more frequent in R-bands or near the border of R- and G-bands. For instance, 78 of the 89 fragile sites (88%) accepted at the Eighth International Workshop on Human Gene Mapping are situated in R-bands (41). A number of fragile sites have been located in R-bands near the border of G-bands (5, 113). Moreover, cancer-associated chromosomal aberrations are also nonrandom, with a limited number of genomic sites consistently involved and frequently associated with cellular oncogenes and fragile sites (67). These observations indicate that R-bands and G/R borders are particularly prone to recombination and raise the question of the role played in these phenomena by the compositional discontinuities at G/R borders and within R bands, as well as by the genomic distribution of Alu sequences, CpG islands and other recombinogenic sequences, such as minisatellites (46).

Chromosomal rearrangements have two important consequences; the activation of oncogenes by strong promoters that have been put upstream of them by the rearrangement (50), and the possibility that some chromosome rearrangements lead, in evolutionary time, to reproductive barriers and speciation (92). It may be speculated that the higher incidence of cancer and rate of speciation (17) shown by warm-blooded relative to cold-blooded vertebrates correlate with the higher propensity to chromosome rearrangements of the former due to the larger number of compositional discontinuities in their chromosomes.

Isochores and Replication/Condensation Timing in the Cell Cycle

As already noted, G-bands and genes located in GC-poor isochores replicate late in the cell cycle, whereas R-bands and genes located in GC-rich isochores replicate early (12, 18, 37). The correspondence of structural (G- and R-) banding and replication (BrdU) banding is, however, only found in warm-blooded vertebrates. Indeed, early and late replication occurs in cold-blooded vertebrates (2, 36, 110) in the absence or poorness of chromosomal banding. Another feature distinguishing G- and R-bands is the condensation timing

during mitosis. This occurs early in the cell cycle for G-bands and late for R-bands. In all probability, early and late condensation also occur in cold-blooded vertebrates in the absence of distinct G- and R-banding.

These findings suggest that replication and condensation timing are correlated with the basic chromomere-interchromomere organization of chromosomes, which is present in cold-blooded vertebrates, and only later in evolution became associated with the compositional differences that arose at the transition between cold-blooded and warm-blooded vertebrates.

That replication timing may be subject to more complex rules under certain circumstances, is indicated by the following observations: although "tissue-specific" genes, which are more frequent in G-bands (37), and in GC-poor isochores (70, 71), generally replicate late, they appear to replicate early in those cell types that express them (27); secondly, the two female X chromosomes have identical G- and R-band patterns, but different replication-band patterns (91).

Isochores, Genomic Distribution of Genes and Codon Usage

As previously indicated, gene concentration is much higher (by a factor of 5–10) in the GC-richest isochores of the human genome than in the other isochores. Moreover, housekeeping genes (including oncogenes) are preferentially distributed in GC-rich isochores (12, 70, 71) and R-bands (37; see also the next section), whereas tissue-specific genes are more abundant in GC-poor isochores and G-bands. These findings indicate that the higher gene concentration in the GC-richest isochores is likely to be underestimated at present, because housekeeping genes are still severely underrepresented in gene banks, and that the genomic distribution of genes is correlated, at least to some extent, with gene function.

The range of GC values (30–100%) in codon third positions of human genes (Figure 5) is almost as wide as that exhibited by the genes of all prokaryotes. This very extended range implies very large differences in codon usage for GC-poor and GC-rich genes of the same genome (8–12, 70, 71). In particular, at the high-GC end of the range, an increasing number (up to 50%) of codons are simply absent. In turn, the wide range of GC values in codon third positions is paralleled by the GC values in first and second positions, although to a more limited extent as would be anticipated. This wide range leads to very significant differences in the frequency of certain amino acids in GC-rich and GC-poor genes. For instance, the ratio of alanine+arginine to serine+lysine (namely, the two amino acids that contribute most to the thermodynamic stability of proteins over the two that do so the least; 4), increases by a factor of four between proteins encoded by the GC-poorest and the GC-richest coding sequences in the human genome (10).

These observations suggest that in warm-blooded vertebrates constitutively

expressed, housekeeping genes tend to be the most biased in codon usage and to direct the synthesis of thermodynamically more stable proteins, compared to tissue-specific genes.

Isochores, CpG Doublets and "CpG Islands"

The distributions of CpG doublets and of CpG islands indicate that the distribution of methylation in the genome of man and other vertebrates is highly nonuniform, a point of interest in view of the functional role of DNA methylation and of the distribution of housekeeping and tissue-specific genes.

The results on CpG islands have an additional functional relevance. In GC-poor isochores, genes are usually endowed with a TATA or a CCAAT box and an upstream control region, whereas in the GC-rich isochores there is no TATA box, but promoters containing properly positioned "G/C boxes" that bind transcription factor Sp1, a protein that activates RNA polymerase II transcription (38, 64, 73, 103). These GC-rich promoters apparently are associated with all genes located in GC-rich isochores, and mainly arose with the appearance of warm-blooded vertebrates (B. Aissani, G. Bernardi & G. Bernardi, in preparation).

ISOCHORES AND GENOME EVOLUTION

Many differences and similarities in compositional patterns were found in vertebrate genomes (8–12, 22, 24, 29, 58, 59, 70, 71, 76, 78, 81, 101, 115); see Figures 1, 2, and 5). A comparison of these patterns sheds new light on genome evolution.

1. The genomes of the vast majority of cold-blooded vertebrates exhibit compositional distributions of both DNA fragments and coding sequences that are narrower and do not reach the GC levels of the GC-rich components of warm-blooded vertebrates and of the coding sequences contained in them. Moreover, a wide spread of GC compositions was found in DNAs from cold-blooded vertebrates (G. Bernardi & G. Bernardi in preparation).

2. In contrast, an overall similarity exists in the compositional patterns of all warm-blooded vertebrates, in spite of the fact that the genomes of birds and mammals not only differ in size by a factor of almost three (the avian genomes being smaller), but also arose separately in evolution. In birds, however, the compositional distributions of both DNA fragments and coding sequences attain higher GC values than in mammals.

3. Finally, two slightly different compositional patterns have been found in mammalian genomes (11, 70, 71, 82, 115; G. Sabeur, J. Filipinski, F. Kadi & G. Bernardi, in preparation). The first one is very widespread in different orders of mammals, and includes the human genome, while the second one appears limited to three families of myomorpha (a sub-order of rodents),

namely murids (rat and mouse), cricetids (hamster) and spalacids (mole rat). The difference essentially consists in a narrower compositional distribution of DNA fragments and coding sequences in the myomorpha compared to other mammals.

These similarities and differences in the compositional patterns of vertebrate genomes define two modes of genome evolution (for a more detailed discussion, see 11). In the conservative mode prevailing in mammals (the relatively minor differences between myomorpha and other mammals are not covered here), the composition of DNA fragments and coding sequences is maintained in spite of a very high degree of nucleotide divergence (which may attain 50% in third codon positions, without correction for multiple hits). This compositional conservation, which can also be observed in banding patterns over a number of chromosomal regions (84), appears to require negative selection operating at the isochore level to eliminate any strong deviation from a presumably functionally optimal composition.

In contrast, in the transitional or shifting mode, parallel compositional changes are seen in both isochores and coding sequences. Investigations on the compositional changes occurring in cold-blooded vertebrates have shown that their extent is not correlated with evolutionary time and may be larger than that of synonymous changes indicating that the molecular clock (116) does not apply during compositional transitions (G. Bernardi & G. Bernardi, in preparation). Two typical compositional transitions are the GC increases that occurred between the genomes of reptiles on the one hand, and birds and mammals on the other (and that were accompanied by the replacement of TATA and CCAAT boxes by GC-rich promoters). These compositional changes are due to a directional fixation of point mutations caused by both negative and positive selection at isochore levels (10–12, 70, 71, 78). Selection appears to be for the higher thermal stabilities of proteins, RNA and DNA, that are required by the higher body temperature of warm-blooded vertebrates. Of course, selection implies functional differences and therefore supports the idea that isochores are functionally relevant structures. Moreover, the compositional relationships between coding and noncoding (particularly intergenic) sequences (Figure 4) indicate that the same compositional constraints apply to both kinds of sequences. The selection pressures underlying such constraints cannot be understood if noncoding sequences are “junk DNA” (74), with no biological function.

CONCLUSIONS: THE PALEOGENOME AND THE NEOGENOME

Isochores represent a new structural level in the organization of the genome of warm-blooded vertebrates that bridges the enormous size gap between the

gene level, with its exon-intron systems and the corresponding regulatory sequences, and the chromosome level, with its banding patterns. These three levels are correlated with each other, since genes match compositionally the isochores in which they are harbored, while GC-poor and GC-rich isochores are DNA segments located in G- and R-bands, respectively.

The investigations that led to the discovery of isochores and of these two correlations (12, 24, 29, 57, 101) have firmly established the existence of differences in the base composition and gene concentration of DNA segments present in G- and R-bands. Moreover, they have revealed that these segments are characterized by strikingly different complexities, the isochores present in G-bands being very close in composition and characterized by a low gene concentration, whereas, the isochores present in R-bands belong to different compositional families, including those of the H3 family that have the highest concentration of genes and CpG islands. The R-band isochores might have a lower DNA compaction, and/or comprise a number of GC-poor isochores; they also appear to be accompanied by different higher order chromatin structures, as judged by DNase sensitivity. These results, as well as the compositional mapping data available so far (see next section), indicate that isochores correspond to a chromosome organization level lower than standard chromosomal bands, possibly to chromomeres and interchromomeres.

Although isochores from the genomes of warm-blooded vertebrates belong to a number of families characterized by large differences in base composition, this is not true for cold-blooded vertebrates. In this case, isochores exist, as shown by hybridization of gene probes and by the fact that the intermolecular compositional heterogeneities of cold-blooded vertebrates are close to those of bacteria in spite of a genome size two-three orders of magnitude higher and of a much larger kinetic complexity (24, 43, 79; G. Bernardi & G. Bernardi, paper in preparation). Isochores from cold-blooded vertebrates are characterized by much smaller differences in composition, which correspond to much weaker banding patterns in metaphase chromosomes. Needless to say, the existence of isochores raises the problem of how these compositionally homogeneous DNA stretches arose in evolution.

Isochores are, however, not only structural units, but also appear to play functional roles. Some of these, like the integration of mobile and viral sequences, recombination and chromosome rearrangements, are well established. In contrast, DNA replication timing, and chromosome condensation timing at mitosis seem rather to be correlated with the chromomere-interchromomere organization of chromosomes, independently of the composition of the corresponding DNA stretches. The observations on the gene distribution in the genome, the relationships of such distribution with gene functions (housekeeping, tissue-specific), with codon usage, and with different kinds of regulatory sequences are also indicative of functional roles for

isochores. In contrast, the possible correlations between isochores and chromatin loops, replicons and transcription units still remain open questions.

Isochores are evolutionary units of vertebrate genomes. Their composition may be conserved in spite of enormous numbers of point mutations, or may undergo dramatic changes after relatively modest numbers of point mutations. In the case of the two independent compositional transitions from cold-blooded vertebrates to mammals and birds, compositional transitions seem to be largely associated with the optimization of genome functions following environmental body temperature changes. Interestingly, these transitions appear to be accompanied by very conspicuous changes in promotor sequences.

To sum up a number of points made in this review, two main compositional compartments can be distinguished in the human genome, and, more generally, in the genomes of warm-blooded vertebrates (see Table 1). The first compartment, the *paleogenome*, is characterized by its similarity to what it was, and still is, in cold-blooded vertebrates: the late-replicating, compositionally homogeneous, GC-poor isochores of early-condensing chromomeres contain relatively rare, GC-poor (largely tissue-specific) genes having

Table 1 The human genome^a

The Paleogenome G-bands, GC-poor isochores	The Neogenome R-bands, GC-rich isochores	Ref.
Chromomeres	Interchromomeres	18
Late replication	Early replication	18
Early condensation	Late condensation	
AT-rich triplex DNA	Z-DNA	16,104
HaeIII digestion	DNase chromomycin A3	55, 90
Abundance of LINES	Abundance of SINES	52, 60, 66, 96, 115
Compositional homogeneity	Compositional heterogeneity	11(b)
Scarcity of genes	Abundance of genes (esp. in H3)	11, 12, 37, 51
GC-poor genes (esp. tissue-specific)	GC-rich genes (esp. housekeeping)	3, 11, 45
Scarcity of CpG islands	Abundance of CpG islands	(c), 106
TATA box promoters	G/C box promoters	26a
DNase I insensitivity	DNase I sensitivity	48
Less frequent recombination	More frequent recombination	

(a) It should be noted (i) that the early replication and late condensation of interchromomeres are neogenome features common to the corresponding compartment of the genome from cold-blooded vertebrates; (ii) that the abundance of genes in the neogenome may derive not only from a higher gene density in the early replicating compartment of cold-blooded vertebrates, but also from gene duplications and/or translocations; (iii) that only the most important properties are listed; secondary effects are not. For example, sites for rare-cutting restriction sites are more abundant where CpG islands are more abundant. Likewise, the abundance of SINES leads to a shorter interspersed pattern of repeats in GC-rich isochores.

(b) K. Gardiner, B. Aissani & G. Bernardi, paper in preparation.

(c) B. Aissani, G. Bernardi & G. Bernardi, paper in preparation.

TATA box promoters (CpG islands are scarce). The second compartment, the *neogenome*, is characterized by the fact that it changed its compositional features compared to what it was in cold-blooded vertebrates. In the *neogenome*, the ancestral, early-replicating, GC-poor isochores of late-condensing interchromomeres were changed into compositionally heterogeneous, GC-rich isochores that contain abundant genes (perhaps including most housekeeping genes) having G/C box promoters (genes and CpG islands are particularly abundant in the GC-richest isochores). The *neogenome/paleogenome* distinction supersedes the first proposal for a bipartite nature of the mammalian genome, that of an ontogenetic genome and a housekeeping genome. That proposal (37) was essentially based on the different distribution of tissue-specific and housekeeping genes, whereas the present proposal is based on the evolutionary history of the compartments.

PERSPECTIVES: COMPOSITIONAL MAPPING

The results discussed in this review suggest a novel experimental approach in human genome research: compositional mapping. Wherever long-range physical maps are available, compositional maps may be constructed by assessing GC levels around landmarks (localized on the physical maps) that can be probed. This simply requires the hybridization of the probes on DNA fractionated according to base composition. If DNA preparations of about 100 Kb in size are used, compositional mapping can define the base composition of DNA stretches of about 200 Kb around the sequence probed. When enough information is obtained, this approach can shed light on the base compositions and the approximate sizes of the isochores making up the genomic regions investigated. As far as chromosome structure is concerned, compositional mapping will provide the equivalent of a high-resolution banding, without the uncertainties of cytogenetics. The compositional map may well correspond to a chromomere-interchromomere map.

This approach has already been successfully tried for a set of 50 unique-sequence probes localized on the long arm of chromosome 21 and has provided a direct demonstration for the compositional homogeneity of G-bands and for the compositional heterogeneity of R-bands q22.1 and q22.3, the highest GC levels (corresponding to the H3 component) being in the telomere-proximal region of q22.3 (K. Gardiner, B. Aissani & G. Bernardi, in preparation). Incidentally, this latter observation agrees with cytogenetical evidence that telomeres almost always correspond to R-bands and that the terminal regions of 20 of them (including that of the long arm of chromosome 21) are the most denaturation-resistant regions of human chromosomes (26). This suggests that H3 may correspond to these denaturation-resistant telomeric regions and to some similar intercalary regions located on chromosomes 11, 19, and 22.

Needless to say, these regions, which contain a considerable fraction of all human genes, should be primary targets in genome sequencing projects. Interestingly, chromosome-specific regions of this kind can be isolated by taking advantage of the different compositional patterns of the genomes of man and myomorpha. Indeed, somatic cell hybrids carrying only one human chromosome (or a part of it) in a rodent background lend themselves to the preparation and cloning of the human GC-richer segments that have no equivalent in the genome of myomorpha and that are characterized by the highest gene concentration.

The modular nature of the isochore organization of the human genome and the properties of isochores from G- and R-bands, respectively, suggest that a detailed knowledge of isochores over a few genomic regions, obviously including the GC-richer ones, should greatly help to understand the organization of the human genome. Once attained, this understanding will not only concern the human genome, or the genomes of warm-blooded vertebrates but eukaryotic genomes in general.

ACKNOWLEDGMENTS

Thanks are due to Toshimichi Ikemura and Alison Stewart for having prompted the writing of this paper; to Brahim Aissani, Giacomo Bernardi, Edwin Crouse, Bernard Dutrillaux, Dusko Ehrlich, Kathleen Gardiner, Christian Gautier, Richard Grantham, Farida Kadi, Julie Korenberg, Ladislav Pivec, Dominique Mouchiroud, Alla Rynditch, Georgette Sabeur, Vittorio Sgaramella, and, especially, to Maxine Singer for comments and criticism.

Literature Cited

1. Abe, K., Wei, J.-F., Wei, F.-S., Hsu, Y.-C., Uehara, H., et al. 1988. Searching for coding sequences in the mammalian genome: the H-2K region of the mouse MHC is replete with genes expressed in embryo. *EMBO J.* 7:3441-49
2. Almeida-Toledo, L. F., Viegas-Pequignot, E., Foresti, F., Toledo Filho, S. A., Dutrillaux, B. 1988. BrdU replication patterns demonstrating chromosome homologies in two species, Genus *Eigenmannia*. *Cytogenet. Cell Genet.* 48:117-20
3. 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.* 14:6345-55
4. Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G. 1979. Thermal stability and protein structure. *Biochemistry* 18:5698-703
5. Aurias, A., Prieur, M., Dutrillaux, B., Lejeune, J. 1978. Systematic analysis of 95 reciprocal translocations of autosomes. *Hum. Genet.* 45:259-82
6. Bernardi, G. 1965. Chromatography of nucleic acids on hydroxyapatite. *Nature* 206:779-83
7. Bernardi, G. 1985. The organization of the vertebrate genome and the problem of the CpG shortage. In *Chemistry, Biochemistry and Biology of DNA Methylation*, ed. G. L. Cantoni, A. Razin, pp. 3-10. New York: Liss
8. Bernardi, G., Bernardi, G. 1985. Codon usage and genome composition. *J. Mol. Evol.* 22:363-65
9. Bernardi, G., Bernardi, G. 1986. The human genome and its evolutionary context. *Cold Spring Harbor Symp. Quant. Biol.* 51:479-87
10. Bernardi, G., Bernardi, G. 1986. Compositional constraints and genome evolution. *J. Mol. Evol.* 24:1-11
11. Bernardi, G., Mouchiroud, D., Gautier, C., Bernardi, G. 1988. Compositional patterns in vertebrate genomes: conservation and change in evolution. *J. Mol. Evol.* 28:7-18
12. Bernardi, G., Olofsson, B., Filipiski, J.,

- Zerial, M., Salinas, J., et al. 1985. The mosaic genome of warm-blooded vertebrates. *Science* 228:953-58
13. Bird, A. 1986. CpG-rich islands and the function of DNA methylation. *Nature* 321:209-13
 14. Britten, R. J., Kohne, D. E. 1968. Repeated sequences in DNA. *Science* 161:529-40
 15. Bünemann, M., Dattagupta, N. 1973. On the binding and specificity of 3,6-bis-(acetato-mercuri-methyl) dioxane to DNAs of different base composition. *Biochim. Biophys. Acta* 331:341-48
 16. Burkholder, G. D., Latimer, L. J. P., Lee, J. S. 1988. Immunofluorescent staining of mammalian nuclei and chromosomes with a monoclonal antibody to triplex DNA. *Chromosoma* 97: 185-92
 17. Bush, G. L., Case, S. M., Wilson, A. C., Patton, J. L. 1977. Rapid speciation and chromosomal evolution in mammals. *Proc. Natl. Acad. Sci. USA* 74: 3942-46
 18. Comings, D. E. 1978. Mechanisms of chromosome banding and implications for chromosome structure. *Annu. Rev. Genet.* 12:25-46
 19. Committee on Mapping and Sequencing the Human Genome. Board on Basic Biology. Commission on Life Sciences. Nat. Res. Council. 1988. *Mapping and Sequencing the Human Genome*. Washington, DC: Natl. Acad. Press
 20. Corneo, G., Ginelli, E., Soave, C., Bernardi, G. 1968. Isolation and characterization of mouse and guinea pig satellite DNA's. *Biochemistry* 7:4373-79
 21. Cortadas, J., Macaya, G., Bernardi, G. 1977. An analysis of the bovine genome by density gradient centrifugation: fractionation in $Cs_2SO_4/3,6$ bis-(acetato-mercurimethyl) dioxane density gradient. *Eur. J. Biochem.* 76:13-19
 22. Cortadas, J., Olofsson, B., Meunier-Rotival, M., Macaya, G., Bernardi, G. 1979. The DNA components of the chicken genome. *Eur. J. Biochem.* 99: 179-86
 23. Cuny, G., Macaya, G., Meunier-Rotival, M., Soriano, P., Bernardi, G. 1978. In *Genetic Engineering*, ed. H. W. Boyer, S. Nicosia, pp. 109-15. Amsterdam, Holland: Elsevier
 24. 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.* 111:227-33
 25. Dodemont, H. J., Soriano, P., Quax, W. J., Ramaeckers, F., Lenstra, J. A., et al. 1982. The genes coding for the cytoskeletal proteins actin and vimentin in warm-blooded vertebrates. *EMBO J.* 1:167-71
 26. Dutrillaux, B. 1973. Nouveau système de marquage chromosomique: les bandes T. *Chromosoma* 41:395-402
 - 26a. Dynan, W. S. 1986. Promoters for housekeeping genes. *Trends Genet.* 2:196-97
 27. Epner, E., Forrestier, W. C., Groudine, M. 1988. Asynchronous DNA replication within the human α -globin gene locus. *Proc. Natl. Acad. Sci. USA* 85: 8081-85
 28. Ferguson-Smith, M. A., Page, B. M. 1973. Pachytene analysis in human reciprocal (10:11) translocation. *J. Med. Genet.* 10:283-86
 29. Filipski, J., Thiery, J. P., Bernardi, G. 1973. An analysis of the bovine genome by Cs_2SO_4/AG^+ density gradient centrifugation. *J. Mol. Biol.* 80:177-97
 30. Fischell-Ghodsian, N., Nicholls, R. D., Higgs, D. R. 1987. Long range genome structure around the human α -globin complex analysed by PFGE. *Nucleic Acids Res.* 15:6197-207
 31. Furst, A., Brown, H., Braunstein, J. D., Schildkraut, C. T. 1981. α -globin sequences are located in a region of early-replicating DNA in murine erythroleukemia cells. *Proc. Natl. Acad. Sci. USA* 78:1023-27
 32. Gardiner, K., Watkins, P., Münke, M., Drabkin, H., Jones, C., Patterson, D. 1988. Partial physical map of human chromosome 21. *Somatic Cell. Mol. Genet.* 14:623-38
 33. Gardiner-Garden, M., Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* 196:261-82
 34. Garel, A., Axel, R. 1976. Selective digestion of transcriptionally active ovalbumin genes from oviduct nuclei. *Proc. Natl. Acad. Sci. USA* 73:3966-70
 35. Gasser, S. M., Laemmli, U. K. 1987. A glimpse at chromosomal order. *Trends Genet.* 3:16-22
 36. Giles, V., Thode, G., Alvarez, M. C. 1988. Early replication bands in two scorpion fishes, *Scorpaena porcus* and *S. notata* (order Scorpaeniformes). *Cytogenet. Cell Genet.* 47:80-83
 37. Goldman, M. A., Holmquist, G. P., Gray, M. C., Caston, L. A., Nag, A. 1984. Replication timing of genes and middle repetitive sequences. *Science* 224:686-92
 38. Goodfellow, P. J., Mondello, C., Darling, S. M., Pym, B., Little, P., Goodfellow, P. N. 1988. Absence of methylation of a CpG-rich region of the 5' end

- of the MIC2 gene on the active X, the inactive X and the Y chromosome. *Proc. Natl. Acad. Sci. USA* 85:5605-9
39. Greilhuber, J. 1977. Why plant chromosomes do not show G-bands. *Theor. Appl. Genet.* 50:121-24
 40. Guttmann, T., Votavova, H., Pivec, L. 1976. Base composition heterogeneity of mammalian DNAs in CsCl₁-netropsin density gradient. *Nucleic Acids Res.* 3:835-45
 41. Hecht, F. 1988. Enigmatic fragile sites on human chromosomes. *Trends Genet.* 4:121
 42. Holmquist, G., Gray, M., Porter, T., Jordan, J. 1982. Characterization of Giemsa dark- and light-band DNA. *Cell* 31:121-29
 43. Hudson, A. P., Cuny, G., Cortadas, J., Haschemeyer, A. E. V., Bernardi, G. 1980. An analysis of fish genomes by density gradient centrifugation. *Eur. J. Biochem.* 112:203-10
 44. Hwu, H. R., Roberts, J. W., Davidson, E. H., Britten, R. J. 1986. Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. *Proc. Natl. Acad. Sci. USA* 83:3875-79
 - 44a. Ikemura, T. 1985. Codon usage and tRNA in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2:13-34
 45. Ikemura, T., Aota, S.-I. 1988. Global variation in G+C content along vertebrate genome DNA. Possible correlation with chromosome band structures. *J. Mol. Biol.* 203:1-13
 46. Jeffreys, A. J., Wilson, V., Thein, S. L. 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* 314:67-73
 47. Kazazian, H. H. Jr., Wong, C., Yousoufian, H., Scott, A. F., Phillips, D. G., Antonarakis, S. E. 1988. Hemophilia-A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in men. *Nature* 332:164-66
 48. Kerem, B.-S., Goitein, R., Diamond, G., Cedar, H., Marcus, M. 1984. Mapping of DNAase I sensitive regions of mitotic chromosomes. *Cell* 38:493-99
 49. Kettmann, R., Meunier-Rotival, M., Cortadas, J., Cuny, G., Ghysdael, J., et al. 1979. Integration of bovine leukemia virus DNA in the bovine genome. *Proc. Natl. Acad. Sci. USA* 76:4822-26
 50. Klein, G. 1983. Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. *Cell* 32:311-15
 51. Korenberg, J., Engels, W. R. 1978. Base ratio, DNA content, and quinine-brightness of human chromosomes. *Proc. Natl. Acad. Sci. USA* 75:3382-86
 52. Korenberg, J. R., Rykowski, M. C. 1988. Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. *Cell* 53:391-400
 53. Kuhn, E. M., Therman, E. 1986. Cytogenetics of Bloom's syndrome. *Can. Genet. Cytogenet.* 22:1-18
 54. Lederberg, J. 1989. The genome project holds promise, but we must look before we leap. *The Scientist* 3:10
 55. Lima-de-Faria, A., Isaksson, M., Olsson, E. 1980. Action of restriction endonucleases on the DNA and chromosomes of *Muntiacus muntjac*. *Hereditas* 92:267-73
 56. Luciani, J. M., Morazzani, M. R., Stahl, A. 1975. Identification of pachytene bivalents in human male meiosis using G-banding technique. *Chromosoma* 52:275-82
 57. Macaya, G., Cortadas, J., Bernardi, G. 1978. An analysis of the bovine genome by density-gradient centrifugation. *Eur. J. Biochem.* 84:179-88
 58. Macaya, G., Thiery, J. P., Bernardi, G. 1976. An approach to the organization of eukaryotic genomes at a macromolecular level. *J. Mol. Biol.* 108:237-54
 59. Macaya, G., Thiery, J. P., Bernardi, G. 1977. In *Molecular Structure of Human Chromosomes*, ed. J. J. Yunis, pp. 35-58. New York: Academic
 60. Manuelidis, L., Ward, D. C. 1984. Chromosomal and nuclear distribution of the Hind III 1.9 Kb human DNA repeat segment. *Chromosoma* 91:28-38
 61. Matassi, G., Montero, L. M., Salinas, J., Bernardi, G. 1989. The isochore organization and the compositional distribution of homologous coding sequences in the nuclear genome of plants. *Nucleic Acids Res.* 17:5273-290
 62. Mayr, E. 1976. *Evolution and the Diversity of Life*, pp. 195, 309. Cambridge, Mass: Harvard Univ. Press
 63. Medrano, L., Bernardi, G., Couturier, J., Dutrillaux, B., Bernardi, G. 1988. Chromosome banding and genome compartmentalization in fishes. *Chromosoma* 96:178-83
 64. Melton, D. W., Konecki, D. S., Brennan, J., Caskey, C. T. 1984. Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proc. Natl. Acad. Sci. USA* 81:2147-51
 65. Meunier-Rotival, M., Cortadas, J., Macaya, G., Bernardi, G. 1979. Isola-

- tion and organization of calf ribosomal DNA. *Nucleic Acids Res.* 6:2109-23
66. Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F., Bernardi, G. 1982. Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. *Proc. Natl. Acad. Sci. USA* 79:355-59
 67. Mitelman, F., Heim, S. 1988. Consistent involvement of only 71 of the 329 chromosomal bands of the human genome in primary neoplasia-associated rearrangements. *Cancer Res.* 48:7115-19
 68. Morgan, W. F., Crossen, P. E. 1977. The frequency and distribution of sister chromatid exchanges in human chromosomes. *Hum. Genet.* 38:271-78
 69. Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., Astrin, S. M. 1988. Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 333:87-90
 70. Mouchiroud, D., Fichant, G., Bernardi, G. 1987. Compositional compartmentalization and gene composition in the genome of vertebrates. *J. Mol. Evol.* 26:198-204
 71. Mouchiroud, D., Gautier, C., Bernardi, G. 1988. The compositional distribution of coding sequences and DNA molecules in humans and murids. *J. Mol. Evol.* 27:311-20
 72. Nelson, M., McClelland, M. 1987. The effect of site-specific methylation on restriction-modification enzymes. *Nucleic Acids Res.* 15:r219-r30
 73. Nur, I., Pascale, E., Furano, A. V. 1988. The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. *Nucleic Acids Res.* 16:9233-51
 74. Ohno, S. 1972. An argument for the genetic simplicity of man and other mammals. *J. Hum. Evol.* 1:651-62
 75. Okada, T. A., Comings, D. E. 1974. Mechanisms of chromosome banding. III. Similarity between G-bands of mitotic chromosomes and chromomeres of mitotic chromosomes. *Chromosoma* 48:65-71
 76. Olofsson, B., Bernardi, G. 1983. Organization of nucleotide sequences in the chicken genome. *Eur. J. Biochem.* 130:241-45
 77. Olofsson, B., Bernardi, G. 1983. The distribution of CR1, an Alu-like family of interspersed repeats in the chicken genome. *Biochim. Biophys. Acta* 740:339-41
 78. Perrin, P., Bernardi, G. 1987. Directional fixation of mutations in vertebrate evolution. *J. Mol. Evol.* 26:301-10
 79. Pizon, V., Cuny, G., Bernardi, G. 1984. Nucleotide sequence organization in the very small genome of a tetraodontid fish *Arothron diadematus*. *Eur. J. Biochem.* 140:25-30
 80. Russell, G. J., Walker, P. M. B., Elton R. A., Subak-Sharpe, J. H. 1976. Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J. Mol. Biol.* 108:1-23
 81. Salinas, J., Matassi, G., Montero, L. M., Bernardi, G. 1988. Compositional compartmentalization and compositional patterns in the nuclear genomes of plants. *Nucleic Acids Res.* 16:4269-85
 82. Salinas, J., Zerial, M., Filipinski, J., Bernardi, G. 1986. Gene distribution and nucleotide sequence organization in the mouse genome. *Eur. J. Biochem.* 160:469-78
 83. Salinas, J., Zerial, M., Filipinski, J., Crépin, M., Bernardi, G. 1987. Non-random distribution of MMTV proviral sequences in the mouse genome. *Nucleic Acids Res.* 15:3009-22
 84. Sawyer, J. R., Hozier, J. C. 1986. High resolution of mouse chromosomes: banding conservation between man and mouse. *Science* 232:1632-35
 85. Schmid, C. W., Hearst, J. E. 1972. Sedimentation equilibrium of DNA samples heterogeneous in density. *Biopolymers* 11:1913-18
 86. Schmid, M., Guttenbach, M. 1988. Evolutionary diversity of reverse (R) fluorescent chromosome bands in vertebrates. *Chromosoma* 97:101-14
 87. Schimke, R. T., ed. 1982. *Gene Amplification*. Cold Spring Harbor, NY: Cold Spring Harbor Lab.
 88. Schwartz, D. C., Cantor, C. R. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37:67-75
 89. Schwartz, D. C., Koval, M. 1989. Conformational dynamics of individual DNA molecules during gel electrophoresis. *Nature* 338:520-22
 90. Schweizer, D. 1977. R-banding produced by DNase I digestion of chromomycin-stained chromosomes. *Chromosoma* 64:117-24
 91. Schweizer, D., Loidl, J., Hamilton, B. 1987. Heterochromatin and the phenomenon of chromosome banding. In *Structure and Function of Eukaryotic Chromosomes*, ed. W. Hennig, pp. 235-54. Heidelberg, Germany: Springer
 92. Seuáñez, H. N. 1979. *The Phylogeny of*

- Human Chromosomes*. Berlin, Germany: Springer-Verlag
93. Singer, M. F. 1982. SINES and LINES: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* 28:433-34
 94. Smith, D. I., Golembieski, W., Gilbert, J. D., Kizyma, L., Miller, J. O. 1987. Over abundance of rare-cutting restriction endonuclease sites in the human genome. *Nucleic Acids Res.* 15:1173-84
 95. Soriano, P., Macaya, G., Bernardi, G. 1981. The major components of the mouse and human genomes. 2. Reassociation kinetics. *Eur. J. Biochem.* 115:235-39
 96. Soriano, P., Meunier-Rotival, M., Bernardi, G. 1983. The distribution of interspersed repeats is non-uniform and conserved in the mouse and human genomes. *Proc. Natl. Acad. Sci. USA* 80:1816-20
 97. Soriano, P., Szabo, P., Bernardi, G. 1982. The scattered distribution of actin genes in the mouse and human genomes. *EMBO J.* 1:579-83
 98. Subak-Sharpe, H., Burk, R. R., Crawford, L. V., Morrison, J. M., Hay, J., Keir, A. M. 1966. An approach to evolutionary relationship of mammalian DNA viruses through analysis of the pattern of nearest neighbor base sequence. *Cold Spring Harbor Symp. Quant. Biol.* 31:737-48
 99. Subak-Sharpe, J. H., Elton, R. A., Russell, G. J. 1974. Evolutionary implications of doublet analysis. *Symp. Soc. Gen. Microbiol.* 24:131-50
 100. Sutherland, G. R., Hecht, F. 1985. *Fragile Sites on Human Chromosomes*. Oxford, England: Oxford Univ. Press
 101. Thiery, J. P., Macaya, G., Bernardi, G. 1976. An analysis of eukaryotic genomes by density gradient centrifugation. *J. Mol. Biol.* 108:219-35
 102. Tikocinski, M. L., Max, E. E. 1984. CG dinucleotide clusters in MHC gene and in 5' demethylated genes. *Nucleic Acids Res.* 12:4385-96
 103. Valerio, D., van der Putten, H., Botteri, F. M., Hoegerbrugge, P. M. 1988. Activity of the adenosine deaminase promoter in transgenic mice. *Nucleic Acids Res.* 16:10083-97
 104. Viegas-Pequignot, E., Derbin, C., Malfoy, B., Taillandier, E., Leng, M., Dutrillaux, B. 1983. Z-DNA immunoactivity in fixed metaphase chromosomes of primates. *Proc. Natl. Acad. Sci. USA* 80:5890-94
 105. Viegas-Pequignot, E., Dutrillaux, B. 1978. Une méthode simple pour obtenir des prophase et des prométaphases. *Ann. Genet.* 21:122-25
 106. Viegas-Pequignot, E., Dutrillaux, B., Thomas, G. 1988. Inactive X chromosome has the highest concentration of unmethylated Hha I sites. *Proc. Natl. Acad. Sci. USA* 85:7657-60
 107. Weinberg, R. A. 1988. To sequence or not to sequence. *Sci. Am.* 259:112
 108. Weiner, A. M., Deininger, P. L., Efstratiadis, A. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55:631-61
 109. Weintraub, H., Groudine, M. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* 93:848-56
 110. Yonenaga-Yassuda, Y., Kasahara, S., Chu, T. H., Rodriguez, M. T. 1988. High-resolution RBG-banding pattern in the genus *Tropidurus* (Sauria, Iguanidae). *Cytogenet. Cell Genet.* 48:68-71
 111. Yunis, J. J. 1976. High resolution of human chromosomes. *Science* 191:1268-70
 112. Yunis, J. J. 1981. Mid-prophase human chromosomes. The attainment of 2,000 bands. *Hum. Genet.* 56:291-98
 113. Yunis, J. J., Soreng, A. L. 1984. Constitutive fragile sites and cancer. *Science* 226:1199-04
 114. Zerial, M., Salinas, J., Filipinski, J., Bernardi, G. 1986. Genomic localization of hepatitis B virus in a human hepatoma cell line. *Nucleic Acids Res.* 14:8373-86
 115. Zerial, M., Salinas, J., Filipinski, J., Bernardi, G. 1986. Gene distribution and nucleotide sequence organization in the human genome. *Eur. J. Biochem.* 160:479-85
 116. ZuckerKandl, E., Pauling, L. 1965. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*, ed. V. Bryson, H. J. Vogel, pp. 97-165. New York: Academic