3

Genome Dynamics and the Generation of Biodiversity

G. Bernardi

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

Summary

Some years ago Wilson and co-workers proposed that the higher rates of karyotypic change and species formation of mammals compared to cold-blooded vertebrates are due to the formation of small demes, as favoured by the social structuring and brain development of the former. Here, evidence is reviewed which indicates that mammals are more prone to karyotypic change and species formation than cold-blooded vertebrates because of their different genome organization. A similar evidence has also recently become available for birds. Although this different organization appears to be a necessary and, in all likelihood, a sufficient condition for the increased rates of karyotypic change and species formation found in mammals, it is still possible that social structuring and brain development may have played an additional accelerating role.

Introduction

Two basic issues in biodiversity are the formation and the extinction of species. As pointed out recently (Benton, 1995) the present-day 5–50 million species were reached by a process of massive diversification, which was not, apparently, a stochastic process, substantial bursts of multiplication of major taxonomic groups coinciding with their invasion of new habitats, which was often associated with the acquisition of new adaptations. This diversification was interrupted by mass extinctions, the largest of which have been identified.

In contrast with species extinction, species formation raises very special problems. The difficulty of producing discontinuities, namely groups (species), from a continuous process (evolution) was realized by Darwin who considered

that speciation was 'the mystery of mysteries' and by Bateson ('the origin and nature of species remains utterly mysterious'). Needless to say, discussions on the nature of speciation assume that species have an objective existence, a point on which not everybody is in agreement. We will accept here Mayr's definition of species as groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups.

It should be briefly recalled here, first, that isolating factors may be prezygotic (e.g. seasonal or habitat isolation) or postzygotic (hybrid inviability and sterility); and, second, that the major speciation models comprise: (i) *allopatry*, physical isolation inevitably leading to evolutionary divergence through natural selection and genetic drift; extensive biogeographical evidence exists in favour of this model; (ii) *sympatry*, involving no physical isolation, but genetic similarity except at the few loci responsible for reproductive isolation and ecological divergence; this model is supported by the cichlids of Lake Victoria; and (iii) *stasipatry*, chromosomal rearrangements reducing fecundity when heterozygous.

We will concentrate on the last model because the view that karyotypic change and species formation are closely linked appears to be supported in the case of the speciation of vertebrates (which will be discussed here) by the work of Wilson and co-workers.

According to White (1968) and Grant (1973), fixation of karyotypic mutations can facilitate species formation and adaptive evolution at the organismal level (see Bush *et al.*, 1977) by acting: (i) as a sterility barrier (stasipatric species formation; White, 1968, 1978), the mutant karyotype functioning at the population level as a cytogenetic reproductive isolating mechanism; (ii) as a regulatory mutation, producing an altered pattern of gene expression that results in an organism with a new and fitter phenotype (see, for example, Zieg *et al.*, 1977); (iii) as a linker of loci that previously were far apart in the genome, thereby creating a particular combination of alleles (Dobzhansky, 1970).

These points were reconsidered by Wilson *et al.* (1974, 1975; Levin and Wilson, 1976; Bush *et al.*, 1977) when they found that cold-blooded vertebrates exhibit a species formation rate which is, on the average, 20% that of mammals, and a karyotypic change rate which is close to 10% that of mammals. Bush *et al.* (1977) proposed that

> The propensity to form small demes is attributable to several factors, one of which may be especially important for understanding how mammals have achieved remarkably high rates of adaptative evolution. This factor is social structuring. If it were not for this social factor, mammals, because of their high dispersal power, might have evolved at the same rate or more slowly than most lower vertebrate.

More recent work, has stressed the importance of brain development in organismic evolution (Wyles *et al.*, 1983; Wilson, 1985).

I will propose here a different explanation for the different karyotypic change rate of cold-blooded vertebrates and mammals. The basic idea,

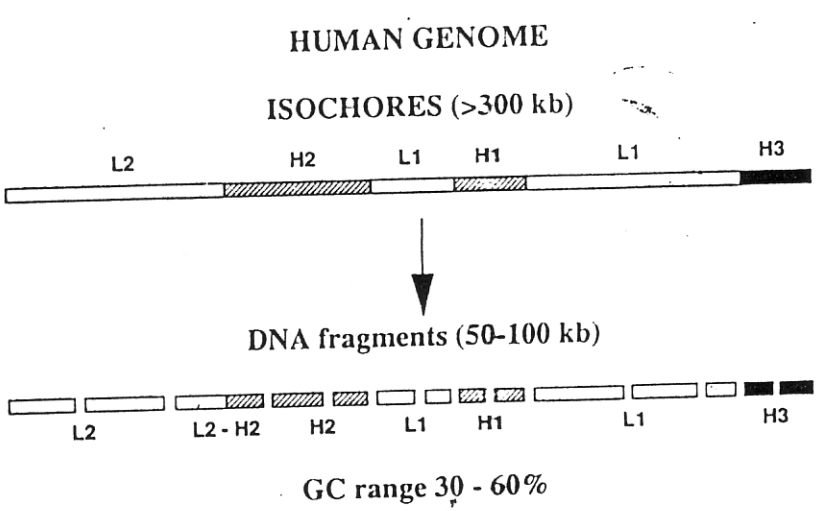


Fig. 3.1. Scheme of the isochore organization of the human genome. This genome, which is a typical mammalian genome, is a mosaic of large (~300 kb) DNA segments, the isochores. These are compositionally homogeneous (above a size of 3 kb) and can be subdivided into a small number of families, GC-poor (L1 and L2), GC-rich (H1 and H2), and very GC-rich (H3). The GC-range of the isochores from the human genome is 30-60% (from Bernardi, 1993b).

first mentioned six years ago (Bernardi, 1989) and further developed later (Bernardi, 1993a), is that the genomes of mammals are more prone to karyotypic changes than those of cold-blooded vertebrates, because of their different organization. The evidence in favour of this explanation will be reviewed. Although the genome organization of mammals appears to be a necessary and, in all likelihood, a sufficient condition for their increased rates of karyotypic change and species formation, the factors discussed by Wilson and co-workers may have played an additional accelerating role.

Genome Organization is Different in Mammals and Cold-blooded Vertebrates

The vertebrate genomes are mosaics of *isochores* (Fig. 3.1), comprising long DNA segments that are homogeneous in base composition (Macaya *et al.*, 1976; Thiery *et al.*, 1976). In warm-blooded vertebrates, isochores (or, more precisely, the 100–200 kb DNA fragments derived from them; see Fig. 3.1) cover a very wide GC range, and attain very high GC values. In the human genome, isochores can be assigned to two GC-poor families (L1 and L2) representing two thirds of the genome, and to three GC-rich families (H1, H2 and H3) forming the remaining one third (Fig. 3.2). In contrast, in cold-blooded vertebrates, isochores are much more uniform in composition and never attain the highest GC levels of the genomes of mammals, as shown by the investigations by Hudson *et al.* (1980) and of Bernardi and Bernardi (1990a,b).

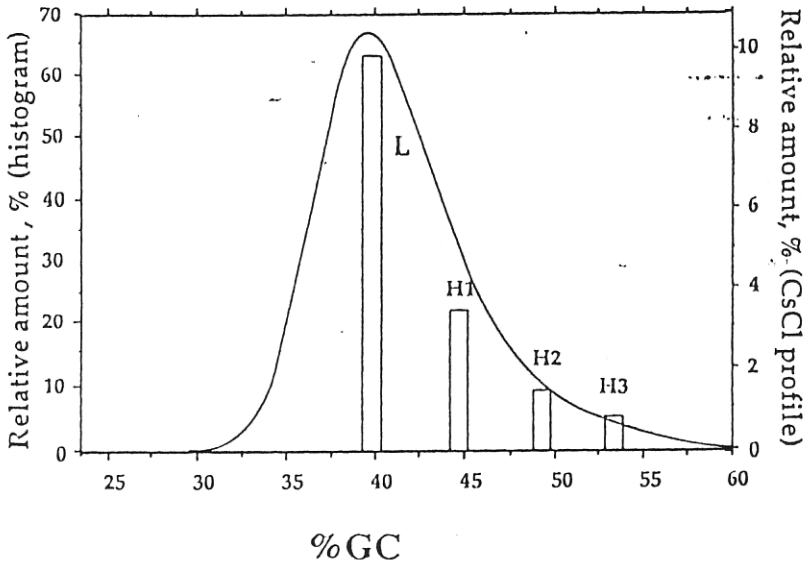


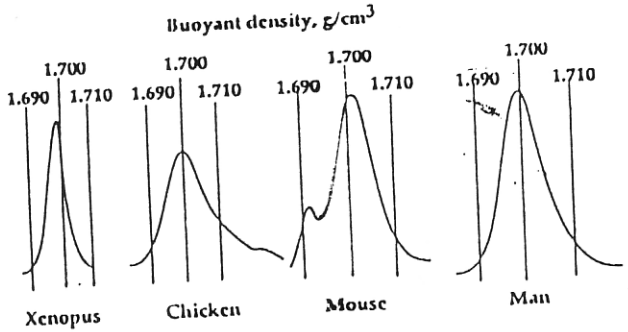
Fig. 3.2. Histogram of the isochore families from the human genome. The relative amounts of major DNA components derived from isochore families L (i.e., L1 + L2), H1, H2, H3 (see Saccone *et al.*, 1993) are superimposed on the CsCl profile of human DNA (from Mouchiroud *et al.*, 1991).

On the other hand, histograms of GC levels of third codon positions of genes from mammals are very highly biased towards very high GC levels. In cold-blooded vertebrates, in contrast, the distribution covers a lower GC range and is more symmetrical (Bernardi *et al.*, 1985, 1988; Bernardi and Bernardi, 1991) (Fig. 3.3). In conclusion, the compositional patterns of mammals and cold-blooded vertebrates are very different, at both the DNA and the coding sequence levels.

Gene distribution is similar in all vertebrates

Compositional correlations (Bernardi *et al.*, 1985) exist between exons (and their codon positions) and isochores (Fig. 3.4), as well as between exons and introns (Aïssani *et al.*, 1991). These correlations concern, therefore, coding and non-coding sequences and are not trivial since coding sequences only make up about 3% of the genome, whereas non-coding sequences correspond to 97% of the genome. The compositional correlations represent a *genomic code* (Bernardi, 1993b). It should be noted that a **universal correlation** (Fig. 3.4) holds among GC levels of codon positions (third positions against first and/or second positions). This is apparently due to compositional constraints working in the same direction (towards GC or AT), although to different extents, on different codon positions, as well as on the isochores.

The compositional correlations between GC_3 (the GC level of third codon



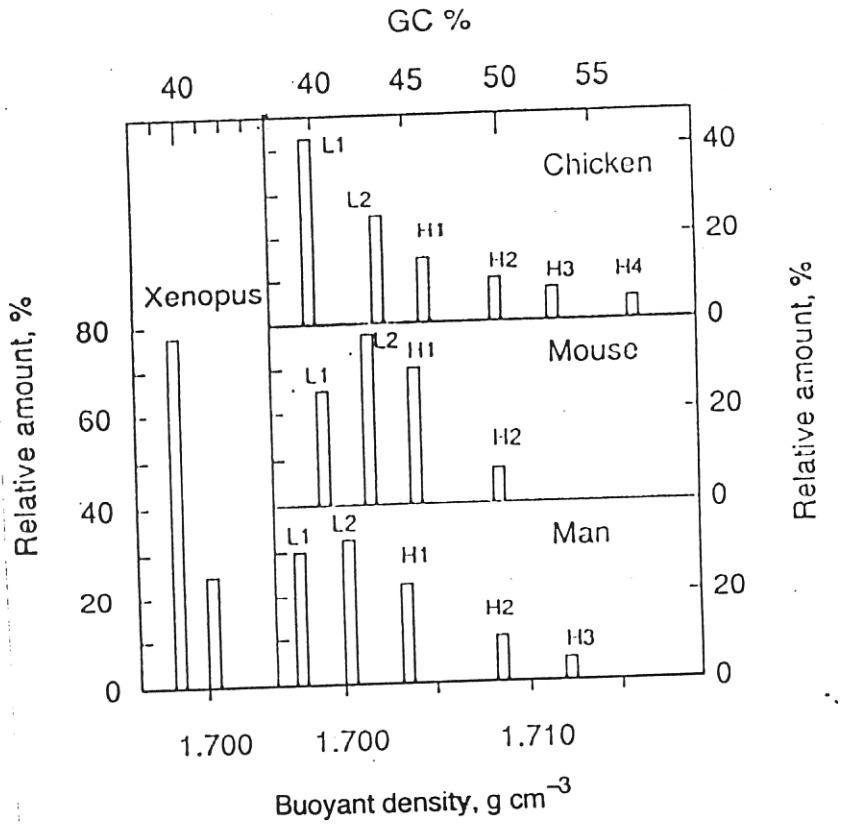


Fig. 3.3. Compositional distribution of third codon positions from vertebrate genes. The number of genes taken into account is indicated. A 2.5% GC window was used. The broken line at 60% GC is shown to provide a reference (from Bernardi, 1993b).

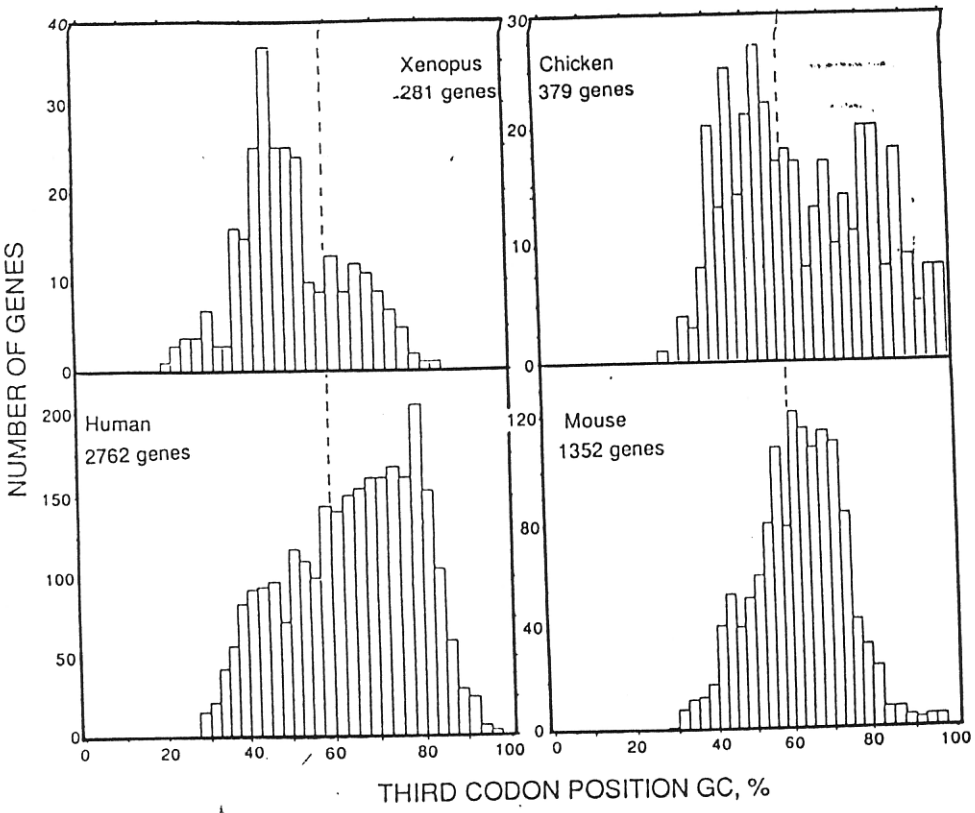


Fig. 3.4. GC levels of third codon positions from human genes are plotted against the GC levels of DNA fractions (dots) or extended sequences (circles) in which the genes are located. The correlation coefficient and slope are indicated. The dash-and-point line is the diagonal line (slope = 1). GC levels of third codon positions would fall on this line if they were identical to GC levels of surrounding DNA. The broken lines indicate a ± 5% GC range around the slope (from Mouchiroud *et al.*, 1991).

positions) and isochore GC have a practical interest in that they allowed us to position the coding sequence histogram of Fig. 3.3 relative to the CsCl profile of Fig. 3.5 and to assess the *gene distribution* in the human genome (Mouchiroud *et al.*, 1991; Bernardi, 1993b). In fact, if one divides the relative number of genes per histogram bar by the corresponding relative amount of DNA, one can see that gene concentration is low and constant in GC-poor isochores, increases with increasing GC in isochore families H1 and H2, and reaches a maximum in isochore family H3, which exhibits at least a 20-fold higher gene concentration compared to GC-poor isochores (Fig. 3.6).

Recent results (mentioned in Bernardi, 1993b) have shown that the gene distribution is similar in all vertebrate genomes in that the isochore family H3 hybridizes preferentially on the GC-richest DNA fractions in all cases.

(B) Plot of GC levels of third codon positions of genes from prokaryotic and eukaryotic genomes are plotted against GC levels of first + second positions. All values are averaged per genome (or per genome compartment, in the case of compositionally compartmentalized genomes) (From D'Onofrio and Bernardi, 1992).

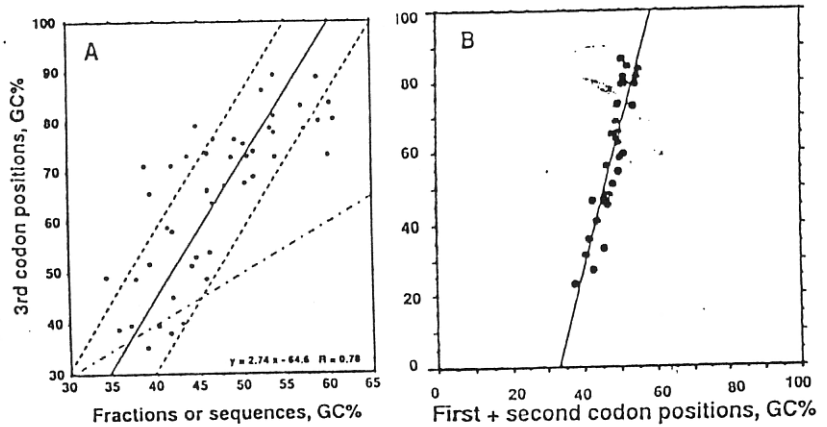


Fig. 3.5. Compositional patterns of vertebrate genomes. Top: CsCl profiles of DNAs from *Xenopus*, chicken, mouse, and man (from Thiery *et al.*, 1976). Bottom: Histograms showing the relative amounts, modal buoyant densities and GC levels of the major DNA components from *Xenopus*, chicken, mouse, and man, as estimated after fractionation of DNA by preparative density gradient in the presence of a sequence-specific DNA ligand (Ag^+ or BAMD; BAMD is bis (acetatomercurimethyl) dioxane). The major DNA components are the families of large DNA fragments (see Fig. 3.1) derived from different isochore families. Satellite and minor DNA components (such as rDNA) are not shown in these histograms (from Bernardi, 1993b).

Chromosome organization is different in cold- and warm-blooded vertebrates

Differences related to those described at the DNA and coding sequence level also exist at the chromosomal level. In human metaphase chromosomes T(elomeric) bands, the subset of R(everse) bands that is most thermal-denaturation resistant, are essentially formed by GC-rich isochores (mainly of the H2 and H3 families). In contrast, R′-bands, namely the R-bands exclusive of T-bands, comprise both GC-rich isochores (of the H1 family) and GC-poor isochores. Finally, G(iemsa) bands are formed almost exclusively by GC-poor isochores (Saccone *et al.*, 1992, 1993; see Fig. 3.7). The difference in GC level between G-bands and T-bands is about 15%. About 20% of genes are present in G-bands and about 80% in R-bands (60% of them in T-bands). The location of a majority of genes in T-bands is of interest in view of the association of telomeres with the nuclear matrix and envelope (de Lange, 1992).

Whereas the organization of metaphase chromosomes just described is essentially valid for all mammals and also for birds, cold-blooded vertebrates are characterized by metaphase chromosomes in which an R banding cannot be elicited (Schmid and Guttenbach, 1988).

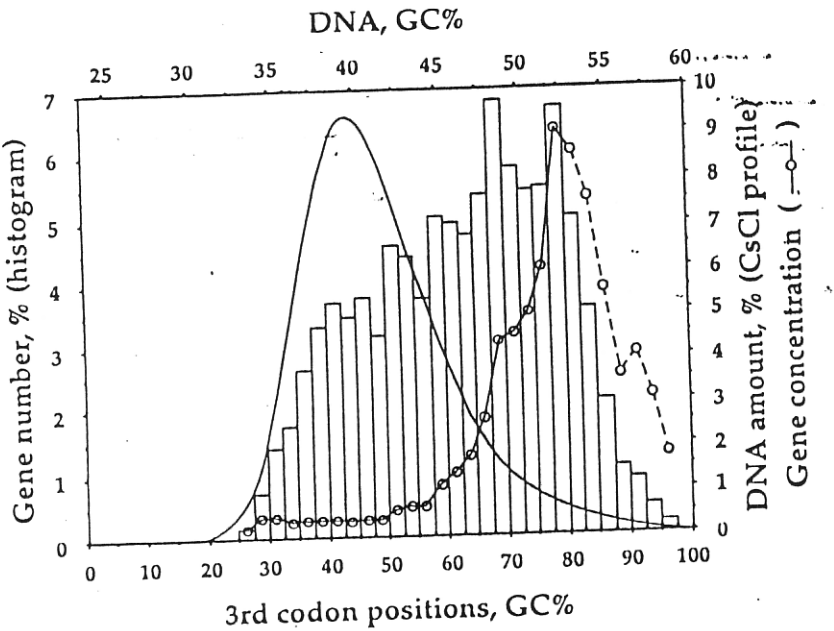


Fig. 3.6. Profile of gene concentration in the human genome as obtained by dividing the relative amounts of genes in each 2.5% GC interval of the histogram by the corresponding relative amounts of DNA deduced from the CsCl profile. The apparent decrease in gene concentration for very high GC values (broken line) is due to the presence of rDNA in that region. The last concentration values are uncertain because they correspond to very low amounts of DNA (from Bernardi, 1993b).

Rearrangements are most frequent at compositional discontinuities

The genomes of warm-blooded vertebrates differ compositionally from those of cold-blooded vertebrates in that about one third of the mammalian genome underwent GC increases, whereas the remaining two-thirds did not. These two compartments were called the neogenome and the paleogenome, respectively (Bernardi, 1989; Fig. 3.8). Such changes took place through the fixation of directional point mutations, as shown by sequence comparisons of homologous genes (Bernardi *et al.*, 1988). The compositional differences in the isochore pattern of warm- compared to cold-blooded vertebrates are paralleled by chromosomal changes, namely the formation of R-bands (and their subset of T-bands).

The crucial point now is that recombinational events in mammals are much more frequent in the Reverse bands and at the Reverse–Giemsa border than in Giemsa bands or than in the genomes of cold-blooded vertebrates.

Detailed information, mainly concerning the human genome, shows that translocation breakpoints are not randomly located on chromosomes (Sutherland and Hecht, 1985; see also Fig. 3.9). Indeed, R-bands and G/R borders are the predominant sites of exchange processes, including

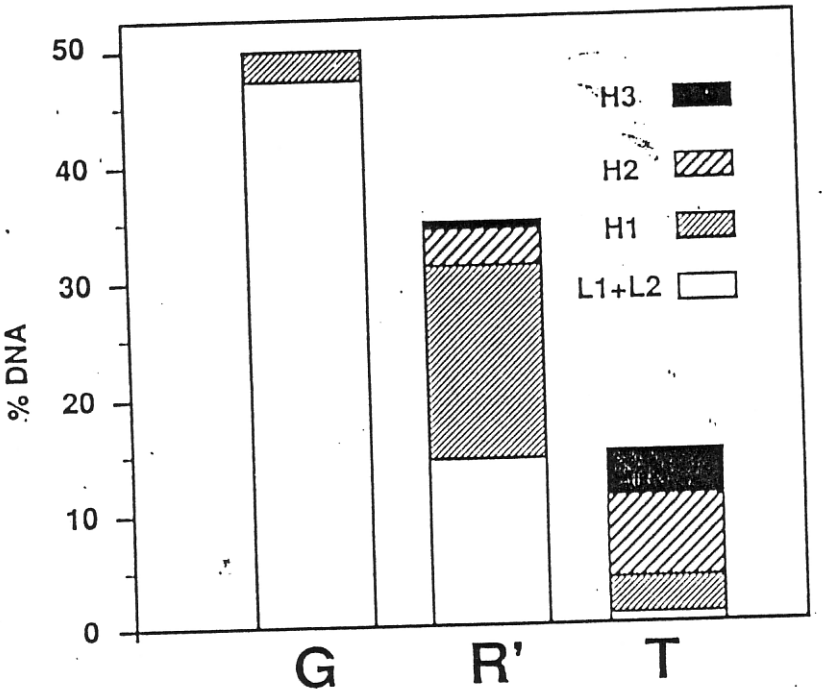


Fig. 3.7. A scheme of the relative amounts of isochore families L1 + L2, H1, H2 and H3 in G-bands, R′-bands and T-bands; R′-bands are R-bands exclusive of T-bands (from Saccone *et al.*, 1993).

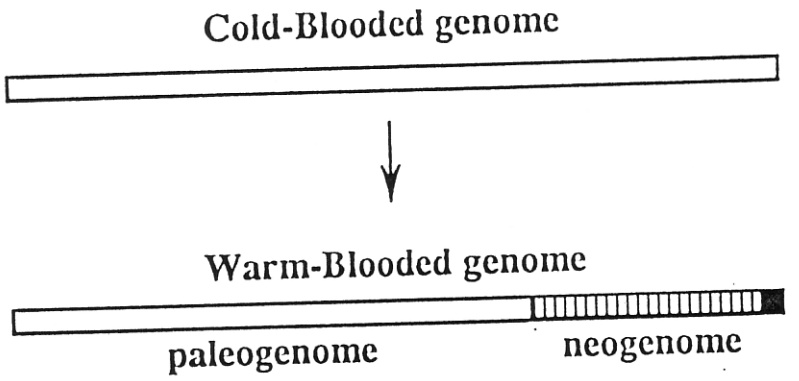


Fig. 3.8. Scheme of the compositional genome transition accompanying the emergence of warm-blooded from cold-blooded vertebrates. The compositionally fairly homogeneous genomes of cold-blooded vertebrates are changed into the compositionally heterogeneous genomes of warm-blooded vertebrates. The latter comprise a paleogenome (corresponding to about two thirds of the genome) which did not undergo any large compositional change and a neogenome (corresponding to the remaining third of the genome, with the GC-richest part only representing 3% of the genome). In the scheme, GC-poor isochores (open bar), GC-rich isochores (hatched bar) and GC-richest isochores (black bar) are represented as three contiguous regions, neglecting the mosaic structure of the isochores in the warm-blooded vertebrate genome (see Fig. 3.1).

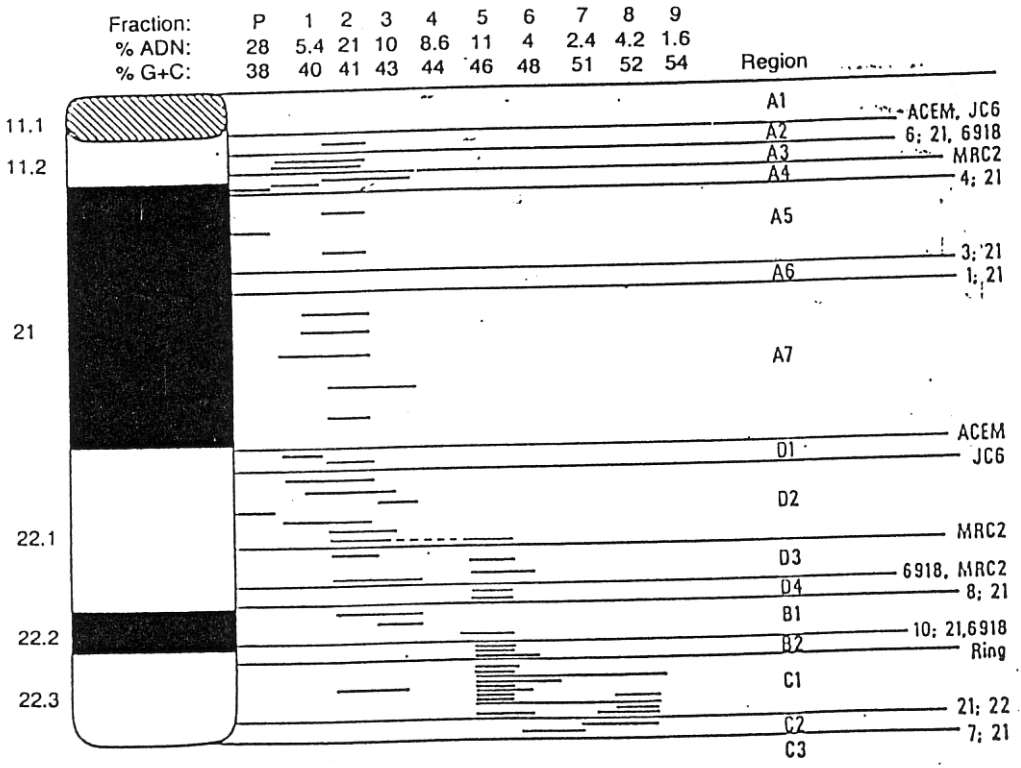


Fig. 3.9. Compositional map of the long arm of human chromosome 21. Long horizontal lines indicate positions of the breakpoints associated with the rearranged chromosomes listed at the right of the figure. Short horizontal lines indicate the compositional DNA fractions hybridizing single-copy probes localized in different Giemsa positive (dark) and Giemsa negative (light) or Reverse bands (from Gardiner *et al.*, 1990).

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 (Morgan and Crossen, 1977; Sutherland and Hecht, 1985; Kuhn and Therman, 1986; Hecht, 1988). Likewise, fragile sites tend to be more frequent in R-bands or near the border of R- and G-bands (Aurias *et al.*, 1978; Yunis and Soreng, 1984; Hecht, 1988). It has been suggested that fragile sites could be the points at which chromosomes break to form rearrangements (Hecht and Hecht, 1984a,b). These observations indicate that R-bands and G/R borders are particularly prone to recombination and suggest that these phenomena are associated with the compositional discontinuities at G/R borders and within R bands. Cancer-associated chromosomal aberrations are also non-random, with a limited number of genomic sites consistently involved and frequently associated with cellular oncogenes and fragile sites (Mitelman and Heim, 1988). Incidentally, chromosomal rearrangements may

have as an important consequence the activation of oncogenes by strong promotors that have been put upstream of them by the rearrangement (Klein, 1983).

At this point, the question should be raised as to why recombination phenomena are so frequent in the neogenome of warm-blooded vertebrates. The high GC level *per se* is unlikely to be responsible because, although rare, genomes which are relatively high in GC (but in a uniform way) are also found among cold- blooded vertebrates (Bernardi and Bernardi, 1990a,b). What seems to matter is, therefore, the presence of compositional discontinuities which exist not only at the borders of G- and R- (or T-) bands, but also within R-and T-bands (Gardiner *et al.*, 1990). The latter are due to the presence of thin G-bands (which can be detected at high resolution; Yunis, 1981) and also to the compositional heterogeneity of the GC-rich isochores present in R- and T-bands.

Concerning the molecular basis for the association between chromosome breakpoints and compositional discontinuities, one possibility is that chromatin structure is more 'open' at compositional discontinuities (as was shown to be the case for CpG islands and the very GC-rich genes which associated with them (Tazi and Bird, 1990; see also Aïssani and Bernardi, 1991a,b)) so allowing recombination to take place, possibly using the many repeated sequences concentrated in those regions. Indeed, the regions which are most susceptible to recombination correspond to high concentrations of Alu sequences (Zerial *et al.*, 1986), CpG islands (Aïssani and Bernardi, 1991a,b) and minisatellites (Jeffreys *et al.*, 1985).

In conclusion, what is proposed here is that the two major compositional shifts which occurred in the genomes of vertebrates, namely those accompanying the emergence of birds and mammals, led to the formation of strongly compositionally compartmentalized genomes which are characterized by a remarkable degree of instability. It should be stressed that, although we do not believe that 'if it were not for social structuring, mammals, because of their high dispersal power, might have evolved at the same rate or more slowly than most cold-blooded vertebrate' (Bush *et al.*, 1977), we can accept that social structuring and brain development may have played an additional accelerating role.

References

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

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

Aïssani, B., D'Onofrio, G., Mouchiroud, D., Gardiner, K., Gautier, C. and Bernardi, G. (1991) The compositional properties of human genes. *Journal of Molecular Evolution* 32, 497–503.

Aurias, A., Prieur, M., Dutrillaux, B. and Lejeune, J. (1978) Systematic analysis of 95 reciprocal translocations of autosomes. *Human Genetics* 45, 259–282.

Benton, M.J. (1995) Diversification and extinction in the history of life. *Science* 268, 52–58.

Bernardi, G. (1989) The isochore organization of the human genome. *Annual Review of Genetics* 23, 637–661.

Bernardi, G. (1933a) Genome organization and species formation in vertebrates. *Journal of Molecular Evolution* 37, 331–337.

Bernardi, G. (1993b) The isochore organization of the human genome and its evolutionary history – a review. *Gene* 135, 57–66.

Bernardi, G. and Bernardi, G. (1990a) Compositional patterns in the nuclear genomes of cold-blooded vertebrates. *Journal of Molecular Evolution* 31, 265–281.

Bernardi, G. and Bernardi, G. (1990b) Compositional transitions in the nuclear genomes of cold-blooded vertebrates. *Journal of Molecular Evolution* 31, 282–293.

Bernardi, G. and Bernardi, G. (1991) Compositional properties of nuclear genes from cold-blooded vertebrates. *Journal of Molecular Evolution* 33, 57–67.

Bernardi, G., Mouchiroud, D., Gautier, C. and Bernardi, G. (1988) Compositional patterns in vertebrate genomes: conservation and change in evolution. *Journal of Molecular Evolution* 28, 7–18.

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

Bush, G.L., Case, S.M., Wilson, A.C. and Patton, J.L. (1977) Rapid species formation and chromosomal evolution in mammals. *Proceedings of the National Academy of Sciences, USA* 74, 3942–3946.

de Lange, T. (1992) Human telomeres are attached to the nuclear matrix. *EMBO Journal* 11, 717–724.

Dobzhansky, T. (1970) *Genetics of the Evolutionary Process*, 3rd edn. Columbia University Press, New York.

Gardiner, K., Aissani, B. and Bernardi, G. (1990) A compositional map of human chromosome 21. *EMBO Journal* 9, 1853–1858.

Grant, V. (1973) *Plant Species Formation*. Columbia University Press, New York.

Hecht, F. (1988) Enigmatic fragile sites on human chromosomes. *Trends in Genetics* 4, 121.

Hecht, F. and Hecht, B.K. (1984a) Fragile sites and chromosome breakpoints in constitutional rearrangements. I. Amniocentesis. *Clinical Genetics* 26, 169–173.

Hecht, F. and Hecht, B.K. (1984b) Fragile sites and chromosome breakpoints in constitutional rearrangements. II. Spontaneous abortions, stillbirths and newborns. *Clinical Genetics* 26, 174–177.

Hudson, A.P., Cuny, G., Cortadas, J. Haschemeyer, A.E.V. and Bernardi, G. (1980) An analysis of fish genomes by density gradient centrifugation. *European Journal of Biochemistry* 112, 203–210.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985), Hypervariable 'minisatellite' regions in human DNA. *Nature* 314, 67–73.

Klein, G. (1983) Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. *Cell* 32, 311–315.

Kuhn, E.M. and Therman, E. (1986) Cytogenetics of Bloom's syndrome. *Cancer Genetics and Cytogenetics* 22, 1–18.

Levin, D.A. and Wilson, A.C. (1976) Rates of evolution in seed plants: net increase in diversity of chromosome numbers and species numbers through time. *Proceedings of the National Academy of Sciences USA* 73, 2086–2090.

Macaya, G., Thiery, J.P. and Bernardi, G. (1976) An approach to the organization

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

of eukaryotic genomes at a macromolecular level. *Journal of Molecular Biology* 108, 237–254.

Mitelman, F. and Heim, S. (1988) Consistent involvement of only 71 of the 329 chromosomal bands of the human genome in primary neoplasia-associated rearrangements. *Cancer Research* 48, 7115–7119.

Morgan, W.F. and Crossen, P.E. (1977) The frequency and distribution of sister chromatid exchanges in human chromosomes. *Human Genetics* 38, 271–278.

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

Saccone, C., De Sario, A., Della Valle, G. and Bernardi, G. (1992) The highest gene concentration in the human genome are in T-bands of metaphase chromosomes. *Proceedings of the National Academy of Sciences USA* 89, 4913–4917.

Saccone, C., De Sario, A., Wiegant, J., Rap, A.K., Della Valle, G. and Bernardi, G. (1993) Correlations between isochores and chromosomal bands in the human genome. *Proceedings of the National Academy of Sciences USA* 90, 11929–11933.

Schmid, M. and Guttenbach, M. (1988) Evolutionary diversity of reverse® fluorescent chromosome bands in vertebrates. *Chromosoma* 97, 110–114.

Sutherland, G.R. and Hecht, F. (1985) *Fragile Sites on Human Chromosomes*. Oxford University Press, Oxford.

Tazi, J, and Bird, A.P. (1990) Alternative chromatin structure at CpG islands. *Cell* 60, 909–920.

Thiery, J.P., Macaya, G. and Bernardi, G. (1976) An analysis of eukaryotic genomes by density gradient centrifugation. *Journal of Molecular Biology* 108, 219–235.

White, M.J.D. (1968) Models of species formation. *Science* 159, 1065–1070.

White, M.J.D. (1978) *Modes of Species Formation*. Freeman & Co, San Francisco.

Wilson, A.C. (1985) The molecular basis of evolution. *Scientific American* 253, 148–157.

Wilson, A.C., Sarich, V.M. and Maxson, L.R. (1974) The importance of gene rearrangement in evolution: evidence from studies on rates of chromosomal, protein, and anatomical evolution. *Proceedings of the National Academy of Sciences USA* 71, 3028–3030.

Wilson, A.C., Busch GL, King MC (1975) Social structuring of mammalian populations and rate of chromosomal evolution. *Proceedings of the National Academy of Sciences USA* 72, 5061–5065.

Wyles, J.S., Kunkel, J.G. and Wilson, A.C. (1983) Birds, behavior and anatomical evolution. *Proceedings of the National Academy of Sciences USA* 80, 4394–4397.

Yunis, J.J. (1981) Mid-prophase human chromosome. The attainment of 2900 bands. *Human Genetics* 56, 291–298.

Yunis, J.J. and Soreng, A.M. (1984) Constitutive fragile sites and cancer. *Science* 226, 1199–1204.

Zerial, M., Salinas, J., Filipski, J. and Bernardi, G, (1986) Gene distribution and nucleotide sequence organization in the human genome. *European Journal of Biochemistry* 160, 479–485.

Zieg, J., Silverman, M., Hilmen, M. and Simon, M. (1977) Recombinational switch for gene expression. *Science* 196, 170–172.