Chromosome Research **9**: 533–539, 2001.

Genes, isochores and bands in human chromosomes 21 and 22

Salvatore Saccone[1,2], Adam Pavliček[3,4], Concetta Federico[2], Jan Pačes[3] & Giorgio Bernardi[4,5,*]
[1] Dipartimento di Protezione e Valorizzazione Agroalimentare, University of Bologna, via F.lli Rosselli 107, 42100 Reggio Emilia, Italy; [2] Dipartimento di Biologia Animale, University of Catania, via Androne 81, 95124 Catania, Italy; [3]Institute of Molecular Genetics – Academy of Sciences of the Czech Republic, Flemingovo 2, Prague, CZ-16637, Czech Republic; [4]Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France; [5]Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy; Tel: +39-081-5833300; Fax: +39-081-2455807; E-mail: bernardi@alpha.szn.it
**Correspondence*

Received 5 March 2001; received in revised form and accepted for publication by Bernard Dutrillaux 8 June 2001

Key words: chromatin condensation, gene density, genome

Abstract

The recently available DNA sequences from chromosomes 21 and 22 enabled us to define the relationships of different band types with isochores and with gene concentration and to compare these relationships with previous results. We showed that chromosomal bands appear as Giemsa or Reverse bands depending not on their absolute GC level, but on the composition GC level relative to those of adjacent contiguous bands. We also demonstrated that the GC-richest, and gene-richest $H3^+$ bands are characterized by a lower DNA compaction compared with the GC-poorest, gene-poorest $L1^+$ bands. Moreover, our results indicate that the human genome contains about 30 000 genes.

Introduction

The human genome is a mosaic of long (> 300 kb), compositionally homogeneous DNA segments, the isochores (Macaya *et al.* 1976), that belong to five families, two GC-poor families, L1 and L2, and three GC-rich families, H1, H2 and H3 that represent about 30%, 33%, 24%, 7.5% and 4–5% of the genome, respectively (for a recent review see Bernardi 2000). Early indications that chromosomal banding may be due to differences in DNA composition (reviewed in Comings 1978) raised the problem of the relationships between G(iemsa) and R(everse) chromosomal bands and isochores. This problem is of special interest because it also concerns the relationship between bands and gene density, since gene concentration increases with increasing GC levels of isochores (Bernardi *et al.* 1985, Mouchiroud *et al.* 1991) reaching up to a 20-fold higher level in H3 compared with L1 isochores (Zoubak *et al.* 1996).

An experimental approach to this problem was the compositional mapping of human genome regions (Bernardi 1989), either by analysis of YACs (yeast artificial chromosomes) carrying human sequences already localized on chromosomes (Gardiner *et al.* 1990, Pilia *et al.* 1993, De Sario *et al.* 1996, De Sario *et al.* 1997) or

Table 1. Classification, relative amounts, and gene densities of bands from chromosomes 21 and 22.

Bands (%)[1]	By staining properties[2]	By isochore content[3]	Gene density[4]	Total gene number[5]
Giemsa (47%)	G1 (13.7%) } G2 (12.6%) }	$L1^+$ (26.3%)	3.0	2500
	G3 (13.1%) } G4 (7.6%) }	$L1^-$ (20.7%)	6.8	4500
Reverse (53%)		$H3^-$ (35.6%)	8.6	9800
	T ($\approx$15%)	$H3^+$ (17.4%)	17.3	9600
Total (100%)				26 500

[1]Relative amounts of different band types were assessed on the basis of band sizes in the 850-band karyotype of Francke (1994).
[2]Here, we call G1–G4 the G bands characterized by four levels of grey (from black to pale grey) as defined by Francke (1994). T bands are the most heat denaturation-resistant R bands of Dutrillaux (1973). Relative amounts of G bands are estimated from Francke (1994), those of T bands from Dutrillaux (1973).
[3]As defined in Saccone *et al.* (1996, 1999) and Federico *et al.* (2000).
[4]This is the average number of genes per Mb found in the bands of chromosomes 21 and 22.
[5]Extrapolated to the total genome, assuming a size of 3200 Mb. If the high compaction of L1 bands (as estimated in the text) is taken into account, the number of genes of L1 bands is 3800 and the total number of genes is 27 800. Lesser corrections apply to other band types but these were not estimated.

by *in-situ* hybridization of isochores on metaphase chromosomes (Saccone *et al.* 1992, 1993, 1996, 1999, Federico *et al.* 2000). The former analysis showed (1) that all isochore families may be represented in a single band (such as Xq28); (2) that a single isochore can cover several bands (as in the 21cen-q21 region) so showing that there is no simple correlation between isochores and bands; and (3) that GC-rich isochores are generally shorter than GC-poor isochores. The second approach, chromosomal compositional mapping, showed high proportions of the GC-richest H3 and of the GC-poorest L1 isochores in a number of R and G bands, respectively. These $H3^+$ and $L1^+$ bands, as they were named (Saccone *et al.* 1996, Federico *et al.* 2000), largely correspond to the most heat denaturation-resistant R bands, namely the T bands of Dutrillaux (1973) and to the two most intensely staining sets of G bands of Francke (1994), called here G1 and G2, respectively (see Table 1). The remaining G and R bands, namely the $L1^-$ (or G3/G4 bands) and the $H3^-$ bands (i.e. the G and R bands not hybridizing L1 and H3 isochores, respectively; see Table 1) are characterized by an intermediate GC composition.

The human karyotype comprises, therefore, at least three compositionally different sets of chromosomal bands. The GC-richest one, the $H3^+$ bands, is not only endowed with the highest concentration of genes, but also with an 'open' chromatin structure, a very high level of transcriptional activity, the highest recombination frequency and the earliest replication timing in the S phase of the cell cycle, whereas the GC-poorest set, the $L1^+$ bands, is characterized by opposite features, and the third set of bands, the $L1^-$ and $H3^-$ bands, by intermediate properties (Bernardi 2000, Federico *et al.* 1998, 2000).

The present analysis of the nucleotide sequences from the long arms of human chromosomes 22 and 21 (Dunham *et al.* 1999, Hattori *et al.* 2000) allowed us to quantify the gene density/GC level relationship, to compare it with the correlation previously determined (Zoubak *et al.* 1996), and to use it in order to estimate the number of genes in the human genome. It also allowed us to draw some new general conclusions on the nature of G and R bands, and to demonstrate a different packing of DNA in the gene-richest and gene-poorest chromosomal bands.

Materials and methods

GC profiles and genes

We used the nucleotide sequences and the identified genes of chromosomes 22 (Dunham *et al.* 1999) and 21 (Hattori *et al.* 2000). GC profiles of these chromosomes were obtained using a window size of 100 kb. 37%, 41%, 46% and 53% GC were taken as the upper values of the L1, L2, H1 and H2 isochore families, respectively.

Correlation of the GC levels and the chromosomal bands

The chromosome ideogram of Francke (1994) at a resolution of 850 bands per haploid genome was used as a reference. In this ideogram, R bands are shown as white bands, G bands are represented as dark bands corresponding to four different levels of grey. The GC profiles of chromosomes 21 and 22 were first aligned on the cytogenetic ideograms (from Francke 1994), then band borders were positioned at sharp changes of GC level (see Results and Discussion); the distribution of genes (considering known and predicted genes, but not pseudogenes) within the above band borders was then assessed, and the gene density per Mb of DNA was estimated for each chromosomal band.

Nomenclature

In this paper, we used the following non-standard terms: G1, G2, G3 and G4 indicate the G bands that show the four levels of grey ranging from black to pale grey in Francke's ideogram (1994). $H3^+$ and $H3^-$ indicate the R bands characterized by high and non-detectable levels of H3 hybridization, respectively (Saccone *et al.* 1996, 1999). $L1^+$ bands indicate those G bands (corresponding to G1/G2 bands; see above) endowed with the highest level of the GC-poorest L1 isochores (Federico *et al.* 2000).

Results and discussion

Chromosomes 21 and 22 exhibit very different band patterns. While chromosome 21 is made of several compositional regions representing all isochore families, chromosome 22 is essentially formed by H2 and H3 isochores, with a sizable contribution of H1 isochores, but contains no L2 isochores (Figure 1). Compositional fluctuations are remarkably high in the GC-richest regions compared with the GC-poorest regions. Moreover, the GC-richest isochores are shorter than the GC-poorest ones. Each band type is characterized by a defined isochore content: $H3^+$ and $L1^+$ bands are almost only composed of H2/H3 isochores and of L1/L2 isochores, respectively. The compositionally intermediate $L1^-$ and $H3^-$ bands are more heterogeneous, containing different proportions of L2, H1 and H2 isochores, with a relatively large amount of H1 isochores. Remarkably, discrete GC levels can be seen in different compositional regions.

Matching of compositional regions with chromosomal bands (at a 850-band resolution; see Figures 1 & 2) was done using three criteria: the average GC level, the discontinuities in GC level at each band border (see below), and the location of genetic markers assigned to chromosomal bands. Our identified band borders were in good agreement with the recent high resolution mapping of the chromosome 22 markers (Kirsh *et al.* 2000), and only two exceptions out of 19 markers were observed. Indeed, STSs stD22S991E.1 and stbK229A8.SP6 were found on contiguous bands compared to their localization on 22q12.1c and 22q13.2b (Kirsh *et al.* 2000), respectively, but very near to the band border. The average GC level of the bands increased from $L1^+$ (G1 + G2) to $L1^-$ (G3 + G4) among the G bands, and from $H3^-$ to $H3^+$ bands among the R bands. Interestingly, while the GC levels of $H3^+$ bands are always higher than those of $L1^+$ bands, in the case of intermediate bands, G bands can be endowed with higher GC levels than those of (non-adjacent) R bands. Indeed, $L1^-$ (G) bands of chromosome 22 are GC-richer than most $H3^-$ (R) bands of chromosome 21. This situation was also found in bands from the same chromosome (compare 21q22.12 and 21q22.2 with 21q11.2 and 21q21.2). The average GC level of a G band is, however, always lower than those of adjacent R bands, and conversely the average GC level of an R band is always higher than those of adjacent G bands. Moreover, the GC levels

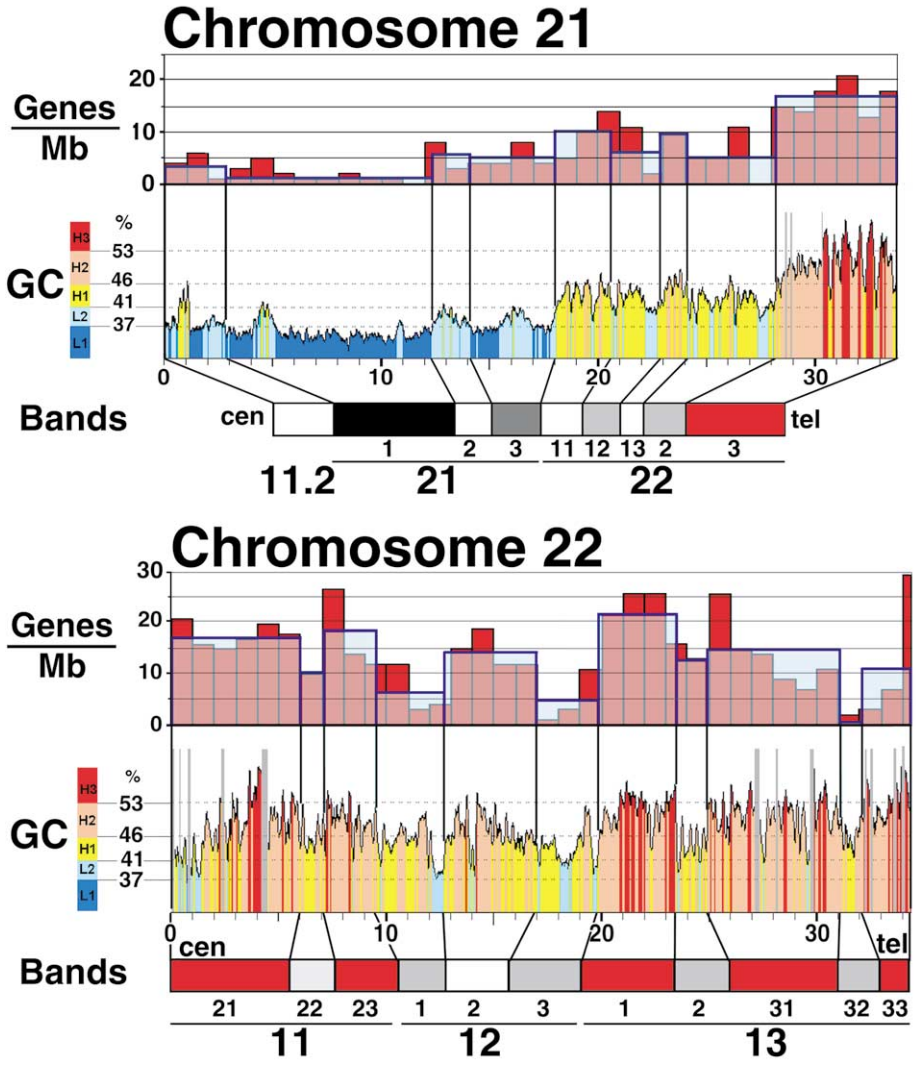

Figure 1. Correlation between chromosomal bands, isochores, and gene concentration of human chromosomes 21 and 22. (Bottom to top) **Bands**: ideogram at a resolution of 850 bands showing the four classes of G bands characterized by different staining intensity (Francke 1994; called here G1 to G4 from black to pale grey); and the two classes of R bands (Saccone *et al.* 1999; $H3^+$, red; $H3^-$, white). The two chromosomes are represented according to their cytogenetic size (Francke 1994). **GC**: GC profiles obtained using a window size of 100 kb; 37%, 41%, 46% and 53% GC were taken as the upper values for the L1, L2, H1 and H2 isochore families, respectively (Zoubak *et al.*, 1996). The grey boxes indicate DNA sequences not yet available. **Genes/Mb**: gene density calculated as number of genes per Mb. The blue bar plot concerns chromosomal bands, the red one 1-Mb segments.

at each band border (over regions of about 300 kb) are always higher on the R side than the G side. Incidentally, this stresses the generality of sharp discontinuities at isochore/band borders first observed by Fukagawa *et al.* (1995). One should, therefore, conclude that the G or R bands are not simply associated with the GC level of given chromosomal regions, but also with the composition of the flanking regions.

Chromosomes 21 and 22 are quite different, not only in their composition, but also in their 'cytogenetic' size. While the DNA amount of the long arms of both chromosomes is comparable (the difference being less than 3%, as judged from

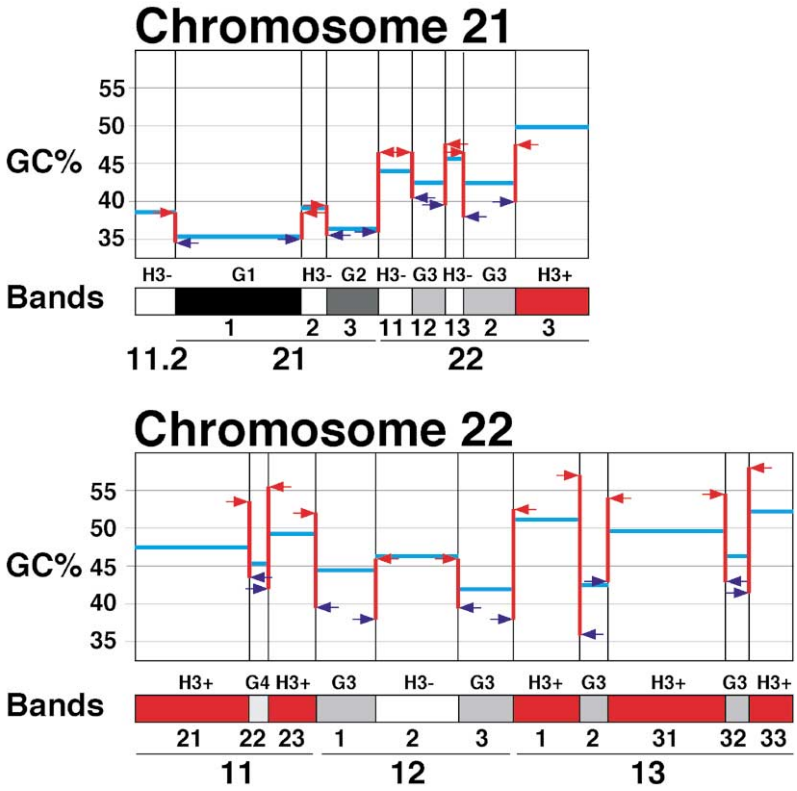

Figure 2. Compositional features of chromosomal bands. (Bottom to top) **Bands**: band ideograms as in Figure 1. G1–G4, $H3^+$, $H3^-$ bands are indicated. **GC%**: average GC level of each chromosomal band (horizontal blue lines), and GC levels observed at band borders (red and blue arrows indicate the GC level on the R and G band side, respectively; vertical red lines indicate the GC difference over 300 kb regions around band borders). Note that all G bands showed lower GC levels than the adjacent R bands, and all R bands showed higher GC levels than the adjacent G bands. These differences were enhanced at band border regions (see above).

sequence data), the cytogenetic size of chromosome 22 was estimated to be about 40% greater than that of chromosome 21 by Francke (1994). This would indicate that DNA is much more compact in chromosome 21 compared with chromosome 22. In turn, this suggests that the size difference between these two chromosomes was overestimated. The same criticism could apply to our results comparing the $L1^+$ band 21q21.1 and the $H3^+$ band 22q13.31. Indeed, these bands show a similar cytogenetic size and a very different DNA sequence size (about 9 Mb and 6 Mb, respectively), indicating a 50% higher level of compactness in the $L1^+$ bands compared with $H3^+$ bands. This comparison could, however, also be criticized because the compared bands belong to two different chromosomes. However, we could estimate about 20% more and 14% less DNA than the average in the $L1^+$ band 21q21.1 and $H3^+$ band 21q22.3 from the same chromosome. Even if quantitative aspects of this different compaction are subject to refinement, especially by analysing larger chromosomes, it is most unlikely that the general qualitative conclusion about differential packing in $H3^+$ and $L1^+$ bands is incorrect. These results provide, therefore, independent additional support, at the band level, that the very active chromatin of the GC-richest and gene-richest chromosomal regions has a more 'open' structure compared with that of the GC-poorest and gene-poorest regions (for a review see Bernardi 2000). Furthermore, at the band level, Yokota *et al.* (1997) analysed three R bands (all of them $H3^+$ bands) and two G bands (one $L1^+$ and one $L1^-$ band) from interphase chromosomes, and observed a particularly striking difference

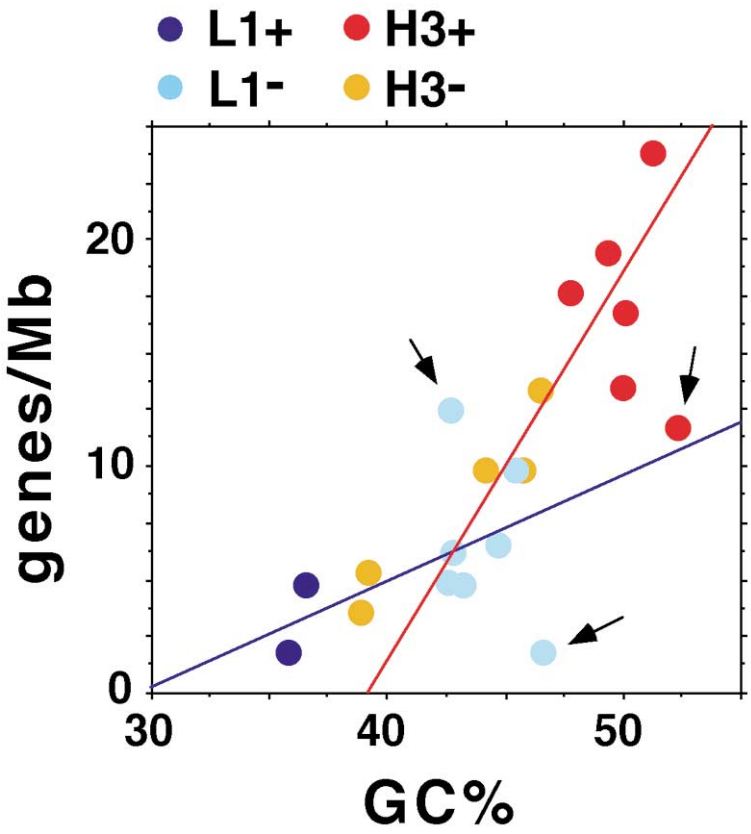

Figure 3. Correlations between average GC levels and the gene densities of chromosomal bands. The average GC level of each band of chromosomes 21 and 22 was plotted against its gene density. The highest and the lowest gene densities were found in $H3^+$ bands and $L1^+$ bands, respectively, as expected. The remaining G and R bands (the $L1^-$ and $H3^-$ bands) showed gene densities that are correlated with their GC level, independently of their cytogenetic band type (G or R). Three points, indicated by the arrows, represent three outliers (two $L1^-$ and one $H3^+$ bands) not taken into consideration when drawing the regression line. Inclusion of these points does not significantly change the lower slope and changes only slightly the higher slope.

between the Xp21.3 ($L1^+$ band) and the Xq28 ($H3^+$ band) without realizing, however, that these bands belonged to the compositionally extreme subsets of G and R bands, respectively.

Figure 1 shows bar plots of gene densities for chromosomal bands, as well as for 1-Mb regions. These data show gene density ratios as high as 20 between $H3^+$ and $L1^+$ bands, and as high as 30 between some 1-Mb regions. The data were used to illustrate the correlation between gene density and GC level of chromosomal bands. This is best described by two regression lines with different slopes that intersect at 42–43% GC (Figure 3), in agreement with previous results obtained on DNA molecules in the 100-kb size range using an independent approach (Zoubak *et al.* 1996).

Finally, the average gene densities found in the bands of chromosomes 21 and 22 can be used to provide an independent approach to estimating the number of genes in the human genome. Indeed, if the four types ($L1^+$, $L1^-$, $H3^-$ and $H3^+$) of bands of chromosomes 21 and 22 are representative, in terms of gene density, of the corresponding bands from the entire karyotype, one can estimate the total number of genes in the human genome as slightly lower than 30 000 (see Table 1). This value is not far from two independent recent assessments, 28 000–34 000 (Ewing & Green 2000), and approximately 35 000

(Roest Crollius *et al.* 2000) based on different approaches, and refines the estimate of 30 500–35 500 obtained by extrapolating gene density of the whole chromosomes 21 and 22 (Dunham 2000) but disagrees with another one, approximately 120 000 (Liang *et al.* 2000). Obviously, our estimate can be further refined by taking into consideration other large chromosomal regions which have just become available and by improving gene prediction methods.

References

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

Bernardi G (2000) Isochores and the evolutionary genomics of vertebrates. *Gene* **241**: 3–17.

Bernardi G, Olofsson B, Filipski J *et al.* (1985) The mosaic genome of warm-blooded vertebrates. *Science* **228**: 953–958.

Comings DE (1978) Mechanisms of chromosome banding and implications for chromosome structure. *Ann Rev Genet* **12**: 25–46.

De Sario A, Geigl EM, Palmieri G *et al.* (1996) A compositional map of human chromosome band Xq28. *Proc Natl Acad Sci USA* **93**: 1298–1302.

De Sario A, Roizès G, Allegre N, Bernardi G (1997) A compositional map of the cen-q21 region of human chromosome 21. *Gene* **194:** 107–113.

Dunham I (2000) The gene guessing game. *Yeast* **17**: 218–224.

Dunham I, Shimizu N, Roe BA *et al.* (1999) The DNA sequence of human chromosome 22. *Nature* **402**: 489–495.

Dutrillaux B (1973) Nouveau système de marquage chromosomique: les bandes T. *Chromosoma* **41**: 395–402.

Ewing B, Green P (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nature Genet* **25**: 232–234.

Federico C, Saccone S, Bernardi G (1998) The gene-richest bands of human chromosomes replicate at the onset of the S-phase. *Cytogenet Cell Genet* **80**: 83–88.

Federico C, Andreozzi L, Saccone S, Bernardi G (2000) Gene density in the Giemsa bands of human chromosomes. *Chromosome Res* **8**(8): 737–746.

Francke W (1994) Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet Cell Genet* **6**: 206–219.

Fukagawa T, Sugaya K, Matsumoto K *et al.* (1995) A boundary of long-range G + C% mosaic domains in the human MHC locus: pseudoautosomal boundary-like sequence exists near the boundary. *Genomics* **25**: 184–191.

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

Hattori M, Fujiyama A, Taylor TD *et al.* (2000) The DNA sequence of human chromosome 21. *Nature* **405**: 311.

Kirsch IR, Green ED, Yonescu R *et al.* (2000) A systematic, high-resolution linkage of the cytogenetic and physical maps of the human genome. *Nature Genet* **24**: 339–340.

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J (2000) Gene Index analysis of the human genome estimates approximately 120,000 genes. *Nature Genet* **25**: 239–240.

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

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

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

Roest Crollius H, Jaillon O, Roest Crollius H *et al.* (2000) Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nature Genet* **25**: 235–238.

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

Saccone S, De Sario A, Wiegant J, Raap AK, Della Valle G, Bernardi G (1993) Correlations between isochores and chromosomal bands in the human genome. *Proc Natl Acad Sci USA* **90**: 11929–11933.

Saccone S, Cacciò S, Kusuda J, Andreozzi L, Bernardi G (1996) Identification of the gene-richest bands in human chromosomes. *Gene* **174**: 85–94.

Saccone S, Federico C, Solovei I, Croquette MF, Della Valle G, Bernardi G (1999) Identification of the gene-richest bands in human prometaphase chromosomes. *Chromosome Res* **7**: 379–386.

Yokota H, Singer MJ, van den Engh GJ, Trask BJ (1997) Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome Res* **5**: 157–166.

Zoubak S, Clay O, Bernardi G (1996) The gene distribution of the human genome. *Gene* **174**: 95–102.