Further Reading

Colin Y, Cherif-Zahar B, Le Van Kim C, *et al.* (1991) Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis. *Blood* **78**: 2747–2752.

Felice AE, Cleek MP, Marino EM, *et al.* (1986) Different ζ-globin gene deletions among black Americans. *Human Genetics* **73**: 221–224.

Ghanem N, Buresi C, Moisan J-P, *et al.* (1989) Deletion, insertion, and restriction site polymorphism of the T-cell receptor gamma variable locus in French, Lebanese, Tunisian, and Black African populations. *Immunogenetics* **30**: 350–360.

Li W-H, Gu Z, Wang H and Nekrutenko A (2001) Evolutionary analyses of the human genome. *Nature* **409**: 847–849.

Mitchell RJ, Howlett S, White NG, *et al.* (1999) Deletion polymorphism in the human *COL1A2* gene: genetic evidence of a non-African population whose descendants spread to all continents. *Human Biology* **71**: 901–914.

Rabbani H, Pan Q, Kondo N, Smith CIE and Hammarström L (1996) Duplications and deletions of the human *IGHC* locus: evolutionary implications. *Immunogenetics* **45**: 136–141.

Rowen L, Koop BF and Hood L (1996) The complete 685-kilobase DNA sequence of the human β T cell receptor locus. *Science* **272**: 1755–1762.

Saitou N and Ueda S (1994) Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Molecular Biology and Evolution* **11**: 504–512.

Teisberg P, Jonassen R, Mevag B, Gedde-Dahl T and Olaisen B (1988) Restriction fragment length polymorphisms of the complement component C4 loci on chromosome 6: studies with emphasis on the determination of gene number. *Annals of Human Genetics* **52**: 77–84.

Weaver DT and DePamphilis ML (1982) Specific sequences in native DNA that arrest synthesis by DNA polymerase α. *Journal of Biological Chemistry* **257**: 2075–2086.

Web Links

Glutathione S-transferase M1 (*GSTM1*); Locus ID: 2944. LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=2944

Glycoprotein, alpha-galactosyltransferase 1 (*GGTA1*); Locus ID: 2681. LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=2681

Opsin 1 (cone pigments), long-wave-sensitive (color blindness, protan) (*OPN1LW*); Locus ID: 5956. LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=5956

Opsin 1 (cone pigments), medium-wave-sensitive (color blindness, deutan) (*OPN1MW*); Locus ID: 2652. LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=2652

T cell receptor beta locus (*TRB@*); Locus ID: 6957. LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=6957

Glutathione S-transferase M1 (*GSTM1*); MIM number: 138350. OMIM: http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?138350

Glycoprotein, alpha-galactosyltransferase 1 (*GGTA1*); MIM number: 104175. OMIM: http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?104175

Opsin 1 (cone pigments), long-wave-sensitive (color blindness, protan) (*OPN1LW*); MIM number: 303900. OMIM: http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?303900

Opsin 1 (cone pigments), medium-wave-sensitive (color blindness, deutan) (*OPN1MW*); MIM number: 303800. OMIM: http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?303800

T cell receptor beta locus (*TRB@*); MIM number: 186930. OMIM: http://www.ncbi.nlm.nih.gov/htbin-post/Omim/dispmim?186930

Gene Distribution on Human Chromosomes

Giorgio Bernardi, *Stazione Zoologica, Napoli, Italy*

Salvatore Saccone, *University of Catania, Catania, Italy*

Intermediate article

Article contents

The distribution of genes in the human genome is strikingly nonuniform. Gene density is very high in GC-rich isochores and very low in GC-poor isochores (isochores are long DNA regions characterized by a fairly homogeneous GC level). Since GC-rich and GC-poor isochores can be localized on human chromosomes, this 'chromosomal compositional mapping' provides information on the distribution of genes on chromosomes.

Introduction

Investigations carried out in our laboratory (for a detailed review, see Bernardi, 2002) have shown that the human genome, like those of all warm-blooded vertebrates, is characterized by a strong compositional heterogeneity. Indeed, these genomes are mosaics of long, compositionally fairly homogeneous regions displaying a broad spectrum of GC levels (GC is the molar ratio of guanine + cytosine in DNA). These regions, called isochores for (compositionally) equal landscapes, range in size from 200 kb to several megabases, and are characterized by very low gene densities in the GC-poor isochore families L1 and L2, which

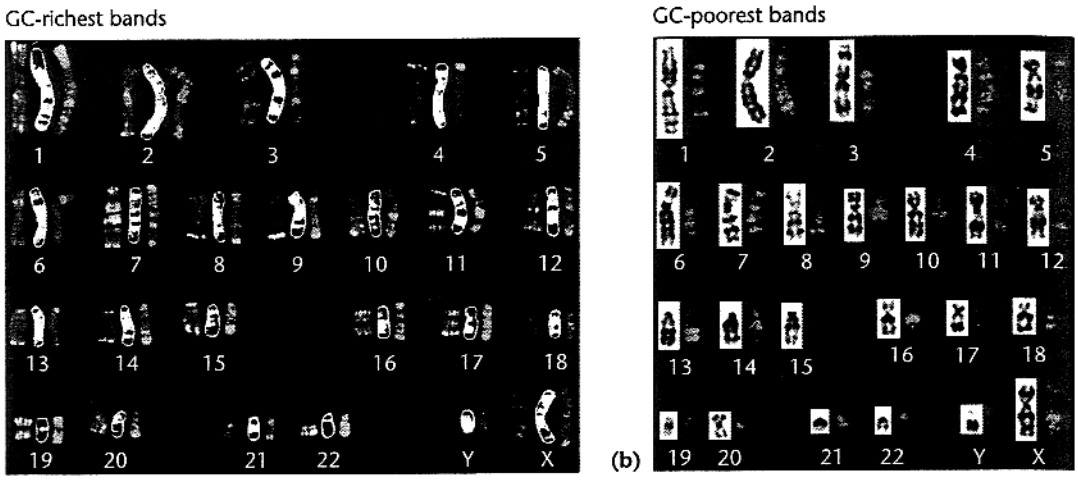

Figure 1 GC-richest and GC-poorest human chromosomal bands. (a) GC-richest bands. From left to right of each chromosome group: T, Chromomycin $A3^+/DAPI^-$ and $H3^+$ bands (from Saccone and Bernardi, 2001). (b) GC-poorest bands. The left and right member of each chromosome pair shows the Giemsa banding and the L1 isochore hybridization respectively (from Federico *et al.*, 2000). Biotinylated H3 and L1 isochores were detected with avidin-FITC and chromosomes were stained with propidium iodide. The H3 and the L1 isochores identify the gene-richest and the gene-poorest chromosomal regions respectively. Note the different localization of the two isochore families, near the telomeres for H3 and more internal for L1 isochores.

form two-thirds of the human genome, and by increasingly higher gene densities in the GC-rich isochore families H1, H2 and H3, which form the remaining third. This strikingly uneven distribution of genes in different isochore families, combined with the possibility of hybridizing compositional DNA fractions *in situ*, provides an approach for investigating the distribution of genes on chromosomes. (***See*** Evolutionary History of the Human Genome; Isochores.)

Gene-richest and Gene-poorest Bands

$H3^+$ and $L1^+$ bands at low resolution

Ten years ago, *in situ* hybridization of compositional fractions of human DNA on metaphase chromosomes (Saccone *et al.*, 1992), under suppression of signals from repeated sequences (Alu, LINES), showed that the H3 isochore family produced the highest concentration of signals on a small number of R(everse), or G(iemsa)-negative bands. These bands largely coincided with two previously defined group of bands:

- the T(elomeric) bands of Dutrillaux (1973), which are the most heat-denaturation-resistant R bands and which are located prevalently in telomeric positions, and
- the chromomycin A3-positive/DAPI-negative bands described by Ambros and Sumner (1987), which are the GC-richest bands of human chromosomes (**Figure 1a**).

Our results were of interest because they showed that these two sets of bands were not due to some satellite or repeated sequence DNA. In fact, they corresponded to the regions of human chromosomes with the highest gene concentration. Moreover, the R bands that contain H3 isochores, the $H3^+$ bands, replicate at the onset of the S phase, whereas the R bands containing no detectable amount of H3 isochores, the $H3^-$ bands, replicate later; G bands replicate even later (Federico *et al.*, 1998, 2000). (***See*** Chromosomal Bands and Sequence Features; Genome Organization.)

If GC-poorest, gene-poorest L1 isochores are hybridized on metaphase chromosomes, the result that is obtained is very similar to that of a G banding, and a large number of hybridized bands (called $L1^+$ bands) coincided with the darkest G bands (**Figure 1b**). The data described so far concern the standard 400 band karyotype.

$H3^+$ and $L1^+$ bands at high resolution

If a higher, 850 band resolution is analyzed by using prometaphase chromosomes, it is clear that the bands displayed at low resolution are compositionally very heterogeneous. This is very evident in the case of the GC-rich chromosomes 19 and 22: at low resolution, these chromosomes are completely covered by the H3 hybridization signals, but, at higher resolution, show bands that do not contain H3 isochores (**Figure 2**). In fact, a chromosomal band at a resolution of 400 bands per haploid genome (R_{400} and G_{400} bands) is generally composed of more than two bands at a

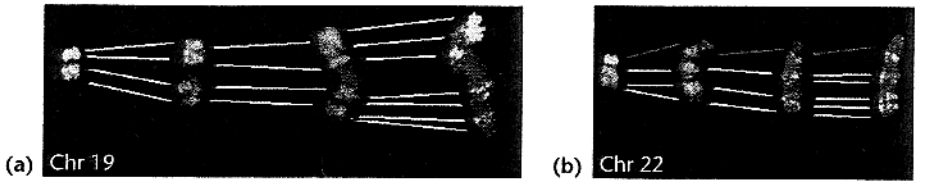

Figure 2 $H3^+$ bands at low and high chromosome resolution. Human chromosomes 19 and 22, at different levels of resolution, hybridized with the biotin-labeled DNA from the H3 isochore family. The hybridized regions were visualized by fluorescein and chromosomes were stained with propidium iodide. Both panels present chromosomes with a band resolution ranging from about 300 to about 850. The bands in chromosome 22 indicate the ribosomal DNA region (modified from Saccone *et al.*, 1999).

resolution of 850 bands per haploid genome (R_{850} and G_{850} bands). For example, the $H3^+$ band 11q13 (**Figure 3**) is composed, at higher resolution, of five bands and more precisely of three R_{850} sub-bands (q13.1, q13.3 and q13.5) and two G_{850} sub-bands (q13.2 and q13.4). Thus, $H3^+$ bands comprise not only R_{850} sub-bands containing H3 isochores, but also G_{850} sub-bands not containing H3 isochores (**Figure 3**). Moreover, a number of other $H3^+$ bands showed that only one of the R_{850} bands derived from each one of them contained H3 isochores, as in the case of the $H3^+$ band 11q23 that yielded only one R_{850} band, 11q23.3, containing H3 isochores (more specifically the distal part of it), whereas the other R_{850} band 11q23.1 and the G_{850} band 11q23.2 do not exhibit H3 isochores (Saccone *et al.*, 1999).

In the case of G banding, at an 850 band resolution, four different types of G bands (on the basis of the degree of staining intensity) were described (Francke, 1994). They were called G1, G2, G3 and G4 bands, from black to pale gray (Federico *et al.*, 2000). *In situ* hybridization of the GC-poorest L1 isochores indicated that this isochore family is only located on a subset of G bands, which largely corresponded to the G1 and G2 bands (Federico *et al.*, 2000; **Figure 3**).

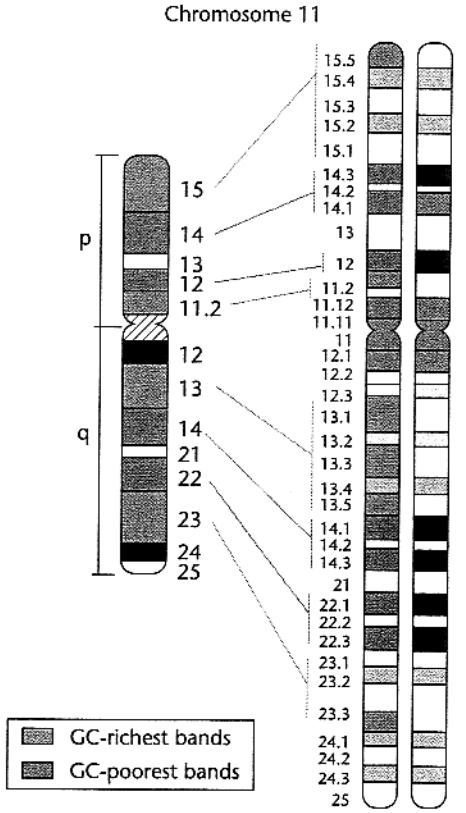

Figure 3 $H3^+$ and $L1^+$ bands on human chromosome 11. Ideogram of human chromosome 11 at 400 (left) and 850 (right) band resolution showing the bands containing the gene-richest H3 and the gene-poorest L1 isochores. At 850 band resolution, the different degrees of darkness of G bands are also shown: note the correspondence among the $L1^+$ bands and the darkest G bands. The R bands not containing H3 isochores are called $H3^-$ bands, and the G bands not containing L1 isochores are called $L1^-$ bands (modified from Saccone *et al.*, 1999; Federico *et al.*, 2000).

Gene Density and Chromosomal Bands

To sum up, human chromosomes comprise four sets of bands: R bands may be very GC-rich (the $H3^+$ bands) or have a modest GC-richness (the $H3^-$ bands); G bands may be very GC-poor (the $L1^+$ bands), or have an intermediate GC level (the $L1^-$ bands; **Figure 3**). Interestingly, the GC-richest bands are preferentially located on telomeric regions and the GC-poorest bands in internal regions of chromosomes (**Figure 1**).

As far as gene concentration is concerned, $H3^+$ bands contain more than 15 genes/Mb, with some subregions higher than 30 genes/Mb, whereas the $L1^+$ bands contain less than five genes/Mb, with some subregions containing one or less gene/Mb (Saccone *et al.*, 2001). This is shown by comparing $L1^+$ and $H3^+$ bands with the compositional maps of chromosomes as obtained from the draft human genome sequence. (***See*** Chromosomes 21 and 22: Gene Density.)

This striking distribution of the GC-richest and the GC-poorest DNA is associated not only with the very different GC levels and gene densities, but also with other features. In fact, the former replicates earlier, is characterized by a more open chromatin, and is more recombinogenic and transcriptionally active compared with the latter. This nonhomogeneous distribution of genes in the chromosomes (and in the nuclei) was also observed in other mammalian and avian species, indicating their conservation in the evolution of chromosomes. (*See* GC-rich Isochores: Origin; Isochores.)

See also

Chromosomal Bands and Sequence Features
Chromosomes 21 and 22: Gene Density
GC-rich Isochores: Origin
Isochores
L Isochore Map: Gene-poor Isochores

References

Ambros PF and Sumner AT (1987) Metaphase bands of human chromosomes, and distinctive properties of telomeric regions. *Cytogenetics and Cell Genetics* **44**: 223–228.

Bernardi G (2002) *Structural and Evolutionary Genomics. Natural Selection in Genome Evolution.* Amsterdam: Elsevier Science (in press).

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

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

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

Francke U (1994) Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenetics and Cell Genetics* **6**: 206–219.

Saccone S and Bernardi G (2001) Human chromosomal banding by *in situ* hybridization of isochores. *Methods in Cell Science* **23**: 7–15.

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

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

Saccone S, Pavliček A, Federico C, Paces J and Bernardi G (2001) Genes, isochores and bands in human chromosomes 21 and 22. *Chromosome Research* **9**: 533–539.

Further Reading

Lander ES, Linton LM, Birren B, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Venter C, *et al.* (2001) The sequence of the human genome. *Science* **291**: 1304–1351.

Gene Duplication: Evolution

Intermediate article

Wen-Hsiung Li, *University of Chicago, Chicago, Illinois, USA*

Article contents

Gene duplication is the major source of genetic novelties, because a duplicate copy may be free to change to a new function. Indeed, genome sequencing and statistical analyses have revealed that each eukaryotic genome contain numerous duplicate genes. Furthermore, there is evidence that a genome duplication occurred in the common ancestor of vertebrates and this might has been important for the diversification and organismal complexity of vertebrates.

Introduction

The evolutionary significance of gene duplication was first recognized by Muller (1935), who suggested that a redundant duplicate gene may acquire divergent mutations and eventually emerge as a new gene. However, few examples of duplicate genes were discovered before the advent of biochemical and molecular biological techniques. The development of protein sequencing methods in the 1950s provided a tool for the study of long-term evolution, and by the late 1950s the α and β chains of hemoglobin were recognized as duplicate genes (Rhinesmith *et al.*, 1958; Braunitzer *et al.*, 1961). Later, isozyme and cytogenetic studies added further evidence for the frequent occurrence of gene duplication in evolution (e.g. Markert, 1964; Harris, 1966; Ritossa *et al.*, 1966). Using data from these various types of studies, Ohno (1970) put forward the view that gene duplication is the only means by which a new gene can arise. Although other means of creating new genes or new functions are now known (see Li, 1997), Ohno's view remains largely valid.

Since the late 1970s, there has been ample evidence that gene duplication is the most important mechanism for generating new genes and new biochemical