GENOMICS 17, 456–462 (1993)

Isochores and CpG Islands in YAC Contigs in Human Xq26.1–qter

GIUSEPPE PILIA,* RANDALL D. LITTLE,* BRAHIM AÏSSANI,† GIORGIO BERNARDI,† AND DAVID SCHLESSINGER*,[1]

*Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110; and †Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2, Place Jussieu, 75005 Paris, France

Received January 28, 1993; revised May 8, 1993

GC levels were assessed at 37 loci across 30 Mb of Xq26.1–qter, a region physically mapped in overlapping yeast artificial chromosome clones. In 8 Mb of R band Xq26, GC is relatively high (up to 44%) in the proximal 4 Mb and relatively low (40–41%) in the distal 4 Mb. Consistently low GC values (38–41%) are observed in G band Xq27. In contrast, further toward the telomere in Xq28, the GC level rises progressively to reach 52% at 2 to 4 Mb from the end of the chromosome; this region is delimited by low GC loci. Across these regions of Xq, the content of rare-cutter restriction enzyme sites containing CpG, including "CpG islands" in the most completely mapped Xq26–27.1 region, is correlated with GC level. Isochore mapping can thus provide one index of putative gene content across mapped regions.

INTRODUCTION

The human genome, like those of warm-blooded vertebrates in general, is a mosaic of large DNA regions (>300 kb on average), the isochores, which are remarkably homogeneous in composition and belong to a small number of families characterized by different GC levels (Bernardi *et al.*, 1985; Bernardi, 1989). In human DNA, two GC-poor isochore families, L1 and L2, and three GC-rich isochore families, H1, H2, and H3, span a range of GC from 30 to 60% (Bernardi, 1989). The L1 and L2 families together represent more than 60% of the genome, whereas the H1, H2, and H3 families correspond to 22, 9 and 5%, respectively.

The distribution of genes in isochore families is strikingly nonuniform (Bernardi *et al.*, 1985), gene concentration in H3 being at least 8 times higher than that in H1 and H2 and at least 16 times higher than that in L1 and L2 (Mouchiroud *et al.*, 1991). Because the GC concentration gradient across isochore families parallels the concentration of genes, isochore maps are especially significant; they have a functional as well as structural relevance and their correlation with cytogenetic bands is of particular interest.

Early studies of metaphase chromosomes indicated that bands staining with Giemsa under standard conditions (G bands) contain GC-poor DNA whereas reverse (R) bands contain GC-rich DNA (Bernardi, 1989). There is, however, no overall correlation with isochores, at least for R bands, as shown by the fact that the ratio of Giemsa to reverse bands is 1:1, whereas that of GC-poor to GC-rich isochores is 2:1 (Bernardi, 1989; Gardiner *et al.*, 1990).

A clarification of the relation between isochores and chromosomal bands or subbands as well as a gene density map can be obtained by constructing "compositional" or isochore maps. To find out where isochores belonging to different families lie along human chromosomes, single-copy sequences that are already localized on a physical map (at least to chromosomal subregions) can be hybridized to human DNA fractionated according to GC content and scored to construct an isochore map (Bernardi, 1989).

Compositional mapping of the long arm of human chromosome 21 (Gardiner *et al.*, 1990) has provided information about GC levels for 53 loci that had been regionally localized. Practically, only GC-poor isochores were detected with probes from G bands, and the most GC-rich isochores were found in the telomeric R band; some low GC regions present along with high GC zones were observed in R bands, and could be due to "thin" G bands (which can be observed at high resolution (Yunis, 1981) or could instead be intrinsic to R bands. Moreover, recent work has shown that the most GC-rich isochore family (H3) in general corresponds (DeSario *et al.*, 1991; Saccone *et al.*, 1992) to the telomeric (T) bands (Dutrillaux, 1973; Ambros and Sumner, 1987).

Instead of using probes with only regional localization, a potentially more informative isochore map can be obtained by examining probes with known location within a yeast artificial chromosome (YAC; Burke *et al.*, 1987) contig. Here we have used probes from contigs spanning 30 Mb in Xq26.1–qter (Schlessinger *et al.*, 1991), a region corresponding to several cytogenetic bands. The result is a detailed correlation of isochores

[1] To whom correspondence should be addressed at Washington University School of Medicine, Department of Molecular Microbiology, 660 South Euclid Avenue, Box 8230, St. Louis, MO 63110. Telephone: (314)362-2744. Fax: (314)362-3203.

0888-7543/93 $5.00

and bands that is further related to the distribution of rare-cutter restriction enzyme sites and CpG islands.

MATERIALS AND METHODS

Clones and probes. Except for two YACs selected from the CGM library of clones from total human DNA (Brownstein *et al.*, 1989; see legend to Fig. 4), YACs were from the X3000.11 hybrid cell, which contains Xq24–qter DNA as its only content of human DNA (Nussbaum *et al.*, 1986). Probes for all the loci tested for GC content are listed in Davies *et al.* (1991) and the Genome Database (GDB) and were kindly supplied by the laboratories of origin or ATCC. Some additional clones from the ends of YACs in Xq26 were also tested for GC content (Fig. 2); they are reported in Little *et al.* (1992) and in the Genome Database. Five additional end-clones of YACs, p477L, p477R, p228R, p414R, p539R, and p512L, from contigs in Xq27 are reported in detail elsewhere (Zucchi *et al.*, in preparation). The probe "ABP" in Fig. 2 is the cDNA for the actin binding protein reported by Gorlin *et al.* (1990). The order of the probes in YAC-based maps across the region Xq26–q28 has been published (Schlessinger *et al.*, 1991), and that in Xq26 is further detailed (Little *et al.*, 1992).

Determination of GC content in isochores containing probes. Total human DNA and fractions of different GC content (described by Gardiner *et al.*, 1990; for other examples of DNA fractionation see Aïssani and Bernardi, 1991; DeSario *et al.*, 1991) were digested with *Eco*RI, electrophoresed, and transferred to a Hybond N^+ membrane (Amersham). Each probe was then labeled and tested for hybridization to the arrayed DNA fractions (Gardiner *et al.*, 1990).

Rare-cutter restriction mapping of YACs. Rare-cutter restriction mapping was carried out using the techniques of total and partial enzyme digestion, electrophoresis, transfer of DNA fragments to Nytran nylon filters, and detection of fragments with left (L) or right (R) YAC vector arms to infer the location of successive sites by indirect end-label mapping, as in Burke *et al.* (1987). Electrophoresis was carried out in a CHEF mapper (Bio-Rad).

RESULTS

Distribution of GC content and isochores across Xq26–qter. To construct a compositional map for Xq26–qter, cloned DNAs mapped at various points in the contig profile (Schlessinger *et al.*, 1991; Little *et al.*, 1992; Davies *et al.*, 1991) were used as probes in Southern blot hybridization against DNA fractions separated by preparative centrifugation in Cs_2SO_4/BAMD density gradients (Bernardi *et al.*, 1985; Bernardi, 1989). This approach establishes the GC level of 100–200 kb surrounding a probed landmark.

A sample analysis for a Factor VIII gene probe is shown in Fig. 1, and GC level is plotted for 37 probes along Xq in Fig. 2. In assessing the results, one must account for the fact that for some probes the GC value is sharply delimited to one compartment of percentage GC, whereas for others, the percentage GC is spread over a range of values. The reason for this is that probes may fall either in a region of very constant GC environment or in a region with some local variation. As a result, regions of higher or lower GC may be close enough to a probe in genomic DNA to be variably included in the random fragments that are separated by density. The

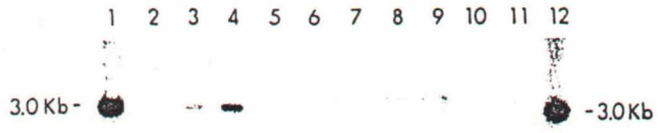

FIG. 1. Isochore mapping of a Factor VIII gene probe. The probe (114.12) was labeled and tested for hybridization to total human DNA and DNA subfractions of different GC content, as described under Materials and Methods and in Gardiner *et al.* (1990). Lanes **1** and **12**, results with total human DNA. Lanes **2–11**, results with isochore fractions obtained by preparative ultracentrifugation in Cs_2SO_4/BAMD at a ligand/nucleotide molar ratio $R_f = 0.14$; the modal GC content of the major component in successive fractions was 38, 40, 41, 43, 44, 46, 48, 51, 52, and 54% (as in Gardiner *et al.*, 1990, and Fig. 2).

fractionations of DNA by density are, however, always quite sharp, and probes at 38–40% and others at 41%, for example, are distinctly nonoverlapping in their values.

The localization of probes in cytogenetic bands is based on standard assignments (Davies *et al.*, 1991). The borders of bands are approximate, estimated from the probe assignments, from *in situ* hybridization experiments with YACs (Montanaro *et al.*, 1991), and from observations of X chromosomes containing translocations [or lacking the terminal portion of the chromosome in cells derived from individuals affected by the Fragile X syndrome (Warren *et al.*, 1990)]. The main results in Fig. 2 can be summarized as follows:

(1) Extensive portions of DNA within cytogenetic bands have relatively constant GC levels.

(2) Xq27, a G band expected to have lower GC content, indeed has an overall GC content lower than that of the R band Xq26. However, there are only weak changes in GC values near the borders of bands defined by cytogenetic staining methods (for example, distal to DXS403), and no changes that might correlate with the subbands Xq26.2 and Xq27.2. Also, comparable GC values were observed across considerable regions of adjacent cytogenetic bands. For example, although a rather wide region centered in the putative region of Xq26.3 is in the distinctly higher 41–44% GC range, at least the more telomeric portion of Xq26 is in the lower 40–41% GC range, similar to many isochores in Xq27.

(3) The GC level in Xq28 increases toward the telomere. The steep rise reaches a peak value well above 50% GC about 3 Mb from the telomere. The gradient is most marked distal to the two locations of the probe for DXS52 and is comparable to that shown earlier (Gardiner *et al.*, 1990) in band q22.3 of chromosome 21. The accentuated most GC-rich peak around the color vision locus (CV in Fig. 2) in the subtelomeric region of Xq28 does not, however, continue to increase monotonically to the telomere, but is flanked by discontinuities in GC level. In the most distal 1.6 Mb, a much lower GC content is seen in the region around Factor VIII. No unique sequences are available in probes closer to the telomere,

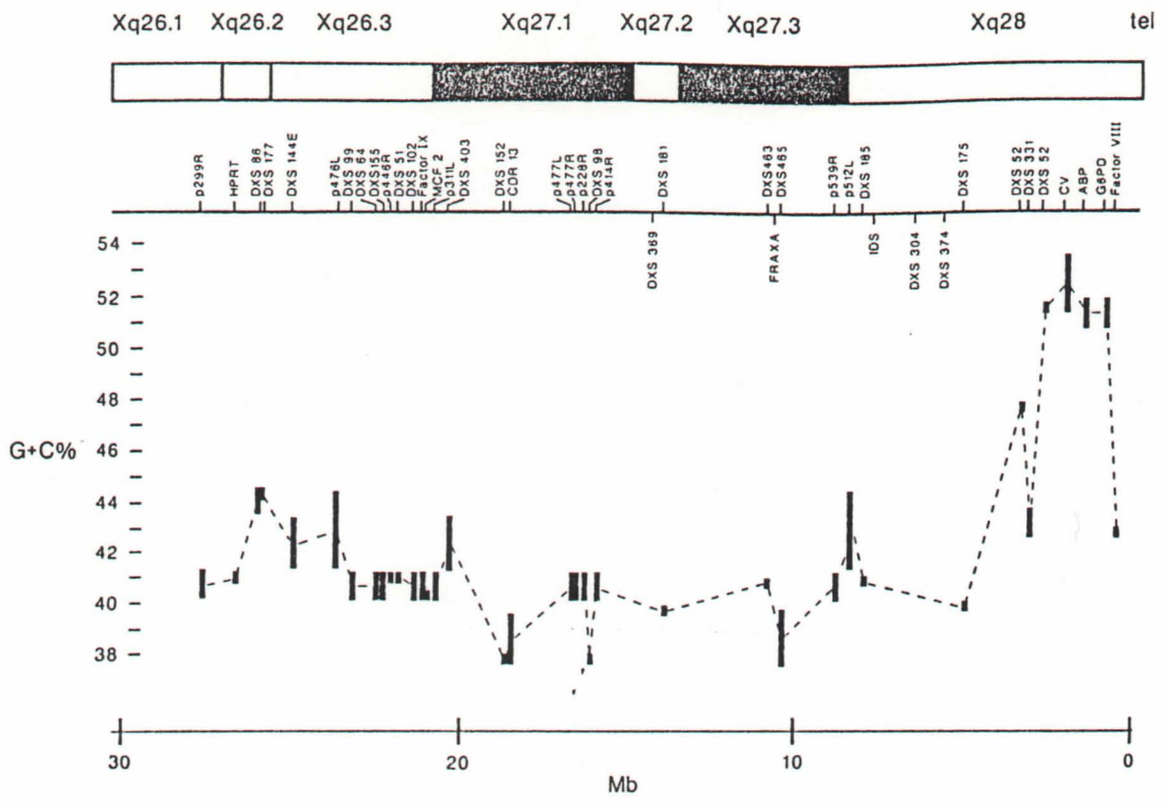

FIG. 2. Placement of 37 probes along a schematic of cytogenetic bands Xq26.1–qter (see text). Values for probes and loci indicated above the bar are plotted according to their GC content (vertical scale), with the length of a filled bar indicating the range of GC values observed in experiments like those of Fig. 1. Distances are shown from the telomere (bottom scale). The positions of additional reference loci are shown below the bar.

but that zone contains at least short sequences with a GC content as high as 85% (Brown *et al.*, 1990).

Rare-cutter mapping and CpG islands in Xq26.1–q27.1. The analysis has been extended to higher resolution for the region spanning Xq26.1–q27 (Fig. 3 and 4) in overlapping YACs (Little *et al.*, 1992). Rare-cutter restriction enzyme mapping of YACs was first used to verify physical distances and to estimate the number of sites for some enzymes that contain CpG in their recognition sites and thereby include potential locations of CpG islands at gene loci (with islands defined according to Bird, 1987; Gardiner-Garden and Frommer, 1987). Such analyses are unambiguous in YACs because the absence of methylation in yeast reveals enzyme sites that may be blocked by methylation in uncloned human DNA. For this phase of the analysis, three rare-cutter enzymes that contain CpG dinucleotides were used, but two are not particularly associated with the CpG islands that define many potential gene loci (Bird, 1987). They would provide an index of sites that might be expected to increase in number with increasing GC level, but would not show any bias toward sites of CpG islands. The other enzyme, *Not*I, would give some estimate of the relative locations of CpG islands, since essentially every *Not*I site is located near an active gene (Bird, 1987).

The region of Xq26.1–q27.1, from the centromeric end of yWXD457 to the telomeric end of yWXD636, contained 64 sites for *Not*I, *Mlu*I, or *Nru*I (Fig. 3). The pattern of cleavage sites was self-consistent in all the YACs of the published contig (Little *et al.*, 1992), and the rare-cutter map with those and other enzymes yielded the same estimate for the overall distance that had earlier been estimated from the probe content of the set of overlapping YACs: about 7.5–8 Mb, or 0.25% of the genome, is included in the region.

As anticipated for a correlation of GC content with rare-cutter sites containing CpG dinucleotides, the centromere proximal 4 Mb, from p299R to just beyond p476L in Fig. 2, which is relatively enriched for GC content, is also enriched for rare-cutter sites for the three enzymes (Fig. 3): 11 of 15 *Not*I; 23 of 25 *Mlu*I; and 18 of 24 *Nru*I sites. The comparable enrichment of *Not*I sites indicated that CpG islands might show a similar correlation.

The mapping was extended to give a further assessment of the internal consistency of YAC structure and to estimate the number and location of potential CpG islands. For this purpose, four additional enzymes that are very often associated with CpG islands were chosen; three of them contain only G and C in their recognition sites (Bird, 1987; Bickmore and Bird, 1992). Figure 4 includes data with the additional enzymes *Sac*II, *Bss*HII, *Sfi*I, and *Eag*I across two sample regions, 2.1 Mb around the HPRT gene and 1.2 Mb around Factor IX. In both cases, YACs were derived from two different sources of

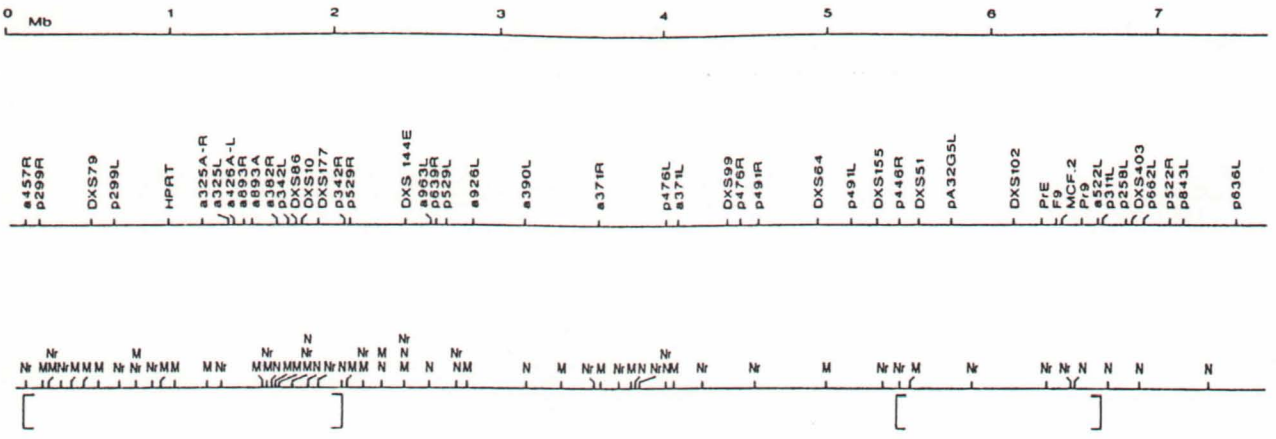

FIG. 3. Rare-cutter restriction maps of the Xq26.1-q27.1 region. The locations of genes and probes for loci are as in Fig. 2. Sites for the rare-cutter enzymes *Mlu*I (M), *Nru*I (Nr), and *Not*I (N) were inferred from pulsed-field gel mapping of YAC DNAs (Burke *et al.*, 1987). Brackets indicate the regions analyzed further for Fig. 4.

DNA (see Fig. 4 legend and Little *et al.*, 1992); and wherever the clones overlap, the same restriction sites are observed at the same distances within experimental error.

Based on these data and the criterion that sites for at least three of the seven enzymes coincide within the resolution of the pulsed-field gels, 20 CpG islands were assigned to the two subregions (these remain potential CpG islands, since there is no evidence about the level of methylation or expression of corresponding genes in cells; see Bickmore and Bird, 1992). The results are in agreement with the expectations (Bird, 1987) that *Not*I

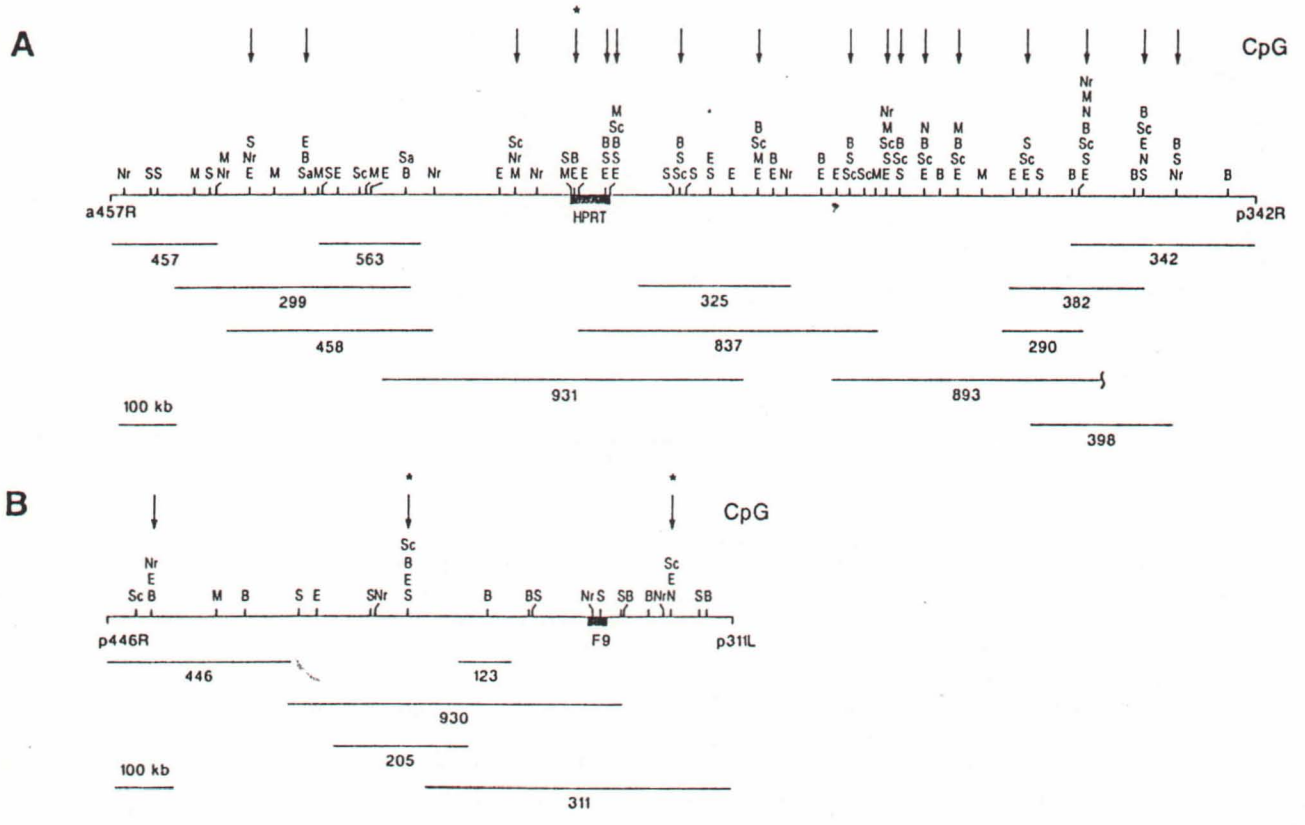

FIG. 4. CpG islands from rare-cutter site coincidence in YACs spanning two subregions of Xq26.1-q27.1. YACs are named by yWXD accession numbers. yWXD931 (containing the HPRT gene; subcontig A) and yWXD930 (containing the F9 gene; subcontig B) were from the CGM library; all other YACs were derived from X3000.11 cells (see Materials and Methods). Locations of CpG islands (indicated by arrows on the level labeled CpG) were scored by the near coincidence of rare-cutter sites for at least three of the seven enzymes used, usually among *Not*I (N), *Bss*HII (B), *Eag*I (E), *Sac*II (Sc), and *Sfi*I (S); *Nru*I (Nr) and *Mlu*I (M) are also shown as in Fig. 3. Three of the CpG islands, marked by asterisks, were previously described (on either side of F9, by Nguyen *et al.*, 1987, and proximal to HPRT, by Wolf and Migeon, 1985). yWXD893 is a chimeric YAC, indicated by the wavy line at its cocloned end (Little *et al.*, 1992). Filled black bars indicate the extent of the F9 gene and the *Mlu*I-*Mlu*I fragment that contains the HPRT gene (Huxley *et al.*, 1991).

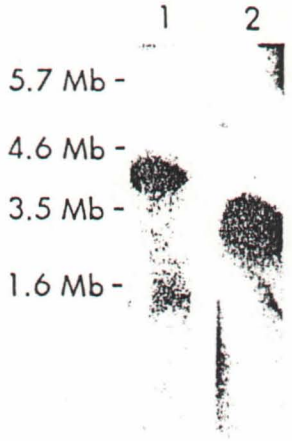

FIG. 5. Pulsed-field gel electrophoretic analysis of the rare-cutter fragments of human HeLa cell DNA that contain the probe sequence for the DXS102 locus (Davies *et al.*, 1991). DNA was digested to completion with each of several enzymes and electrophoresed in parallel with *Schizosaccharomyces pombe* chromosomes as a size marker. Electrophoresis was carried out in TAE buffer and 0.8% agarose at 14°C on a CHEF mapper (Bio-Rad) with the manufacturer's regimen to separate DNA of 400 to 6000 kb and markers including *S. pombe* chromosomes and the largest (1.6 Mb) *Saccharomyces cerevisiae* chromosome. DNA was then transferred to a Nytran nylon membrane and tested for hybridization to a radiolabeled probe for DXS102. The corresponding restriction fragments are the largest found for this region for *Not*I (lane 1) and *Mlu*I (lane 2).

sites usually fall into CpG islands (in these regions, 4 of 4), and that *Sac*II, *Bss*HII, and *Eag*I are also indicators of choice for CpG islands (Bird, 1987). Overall, 17 islands (1/120 kb) were found in the GC-rich proximal and 3 (1/360 kb) in the GC-poorer distal portion. Thus, in agreement with Aïssani and Bernardi (1991), CpG islands correlated as well with GC content as did the total content of *Mlu*I, *Nru*I, and *Not*I sites.

The data from overlapping YACs can also be compared to results with digests of uncloned human DNA. According to Fig. 3, probe sequences should be found in restriction fragments of predictable size. The map has in general been verified for the regions tested. For example, in the region around Factor IX, the enzyme sites indicated are in agreement with those detected for uncloned DNA from another human donor (Nguyen *et al.*, 1987).

Some fragments were larger than expected, consistent with the expected DNA methylation at specific sites in various cultured cells (as shown, for example, for Xq27.3–qter in the study of Poustka *et al.*, 1991, and for the HPRT gene region by Wolf and Migeon, 1985). The largest *Not*I and *Mlu*I fragments detected from the region are shown in Fig. 5, as seen in control DNA extracted from HeLa cells. The largest *Not*I fragment expected from the region, about 2.6 Mb from DXS99 through Factor IX to the mcf2 gene, was included in an even larger fragment of about 4 Mb seen in hybridization of DNA that had been digested, fractionated by pulsed-field gel electrophoresis, and probed with DNA for the locus DXS102 (Fig. 4, lane 1). The same probe detected a 3.5-Mb *Mlu*I fragment (Fig. 4, lane 2) and was also seen in an expected 400-kb *Nru*I fragment (data not shown). Further studies would be required to determine which sites are selectively methylated.

DISCUSSION

Detection of high-GC regions in mapped portions of the genome. In an analysis of the long arm of chromosome 21 (Gardiner *et al.*, 1990), 63% of the telomeric band 22.3 was accounted for in terms of composition, but much lower fractions (about 10%) of other bands were analyzed, and probes were only regionally ordered. In this study, in contrast, only 10 to 20% of the Xq26–qter region was accounted for compositionally, but the order of the landmarks tested and the distances between them were more precise. The present results with probes from a contig-based map of distal Xq agree with the results for chromosome 21 in three important respects:

(1) G bands contained GC-poor isochores.

(2) R bands were heterogeneous in GC level, ranging from portions as low as G-bands to more moderate levels.

(3) DNA of the highest average GC content was localized near the telomere. This general result is not surprising, since the most GC-rich DNA has been shown to correspond to T bands (DeSario *et al.*, 1991; Saccone *et al.*, 1992). The observed most GC-rich region in Xq28 is remarkable, however, in the sense that it does not correspond to a T band (Dutrillaux, 1973; Ambros and Sumner, 1987) nor is it labeled with H3 DNA as a probe by *in situ* hybridization (Saccone *et al.*, 1992). This is likely due to the averaging of hybridizing signals over 0.5% (15 Mb) stretches of DNA (Saccone *et al.*, 1992). Thus, the small proportion of very GC-rich DNA in Xq28 is apparently revealed with greater sensitivity by the direct isochore analysis. The current analysis has employed probes only at intervals of about a million basepairs, and each probe defines the GC level of about 200 kb of surrounding DNA. Thus, the analysis could be refined much further, to the level of neighboring isochores, by studying many additional probes (as in the analyses of the CFTR gene region by Krane *et al.*, 1991, and of the DMD region by Bettecken *et al.*, 1992). In any case, the results indicate that such T bands may prove to be even more widespread than has been suggested by cytogenetic analysis (Dutrillaux, 1973; Ambros and Sumner, 1987).

Incidence of CpG islands and genes across Xq26–qter: Some implications. No assay of gene content has been carried out, but as discussed below, the results are in full accord with the relative distribution of CpG islands and genes expected from isochore analyses (Bernardi *et al.*, 1985; Bernardi, 1989; Mouchiroud *et al.*, 1991; Aïssani

and Bernardi, 1991) and available pulsed-field gel electrophoretic analyses.

The results indicate a correlation among the contents of GC, of rare-cutter sites containing CpG dinucleotides, and of CpG islands. The distal portion of Xq26, which is low in GC (40.5%), contained fewer rare-cutter sites in its 4 Mb compared to the more proximal 4 Mb, which is higher in GC (43%). For example, summing the results for *Not*I, *Mlu*I, and *Nru*I, a total of 43 of 57 sites, or about threefold more, were in the higher GC half of the region. In the test subregions analyzed further with additional rare-cutter enzymes, the more GC-rich zone again contained a density of CpG islands about threefold greater (1/120 kb compared to 1/360 kb).

Published and ongoing work shows that Xq28 exhibits a trend in GC level comparable to that in CpG island content. Proximal portions around IDS (Palmieri *et al.*, 1992) and DXS304 (Palmieri *et al.*, 1993), which contain lower GC, have a CpG island roughly every 250 kb, whereas in distal Xq, in the region of high GC (52%) between color vision and G6PD, there are up to eightfold more CpG islands, with one every 30 to 50 kb (Palmieri *et al.*, 1992; unpublished data). These results are consistent with the observations of Poustka *et al.* (1992) and Dietrich *et al.* (1992) that rare-cutter sites cluster in a region corresponding to the zone of highest GC as shown in Fig. 2. Also consistent with this inference is a study of cosmids from Xq24–q28 containing CpG islands. In that work, 56 such cosmids came from Xq28 and only 6 from Xq24–q27 (Kaneko *et al.*, 1992), and Maestrini *et al.* (1990) report a correspondingly high yield of cloned CpG islands from Xq28.

An index of the levels of CpG islands based on the GC content of a region is no substitute for the more incisive determination of CpG islands by restriction mapping, which also localizes the genes. Also, as shown by the comparisons of Figs. 2 and 3, the observed increase in CpG island content was much more than proportional to increases in GC. However, a preliminary characterization of possible CpG island density across a region can be achieved with a small set of hybridization blots like those in Fig. 1, whereas the rare-cutter analyses of Fig. 4 require about 70 pulsed-field gel analyses. In addition, GC level is a particularly strong index for regions with very high levels of CpG islands, like the one observed in distal Xq28. In that region, the highest GC levels correspond to H3 isochores, which have been assessed to contain about $\frac{1}{4}$ of human genes in about 100 Mb of DNA (Mouchiroud *et al.*, 1991). Based on current estimates of about 50–100,000 human genes, 2 Mb of DNA surrounding the color vision locus might then encode hundreds of genes—a concentration that could approach the gene density now established for the nematode *Caenorhabditis elegans* (Sulston *et al.*, 1992).

The results are in accord with the findings of genetic linkage of a large number of inherited diseases to Xq28 and distal to DXS52 for most of those disease genes that have been well-mapped (Davies *et al.*, 1991; Fig. 2). Those findings have been puzzling, for it seemed improbable that genes likely to be modified in inherited disease states would somehow cluster near the telomere. Instead, the clustering can be simply related to the higher GC content, with its accompanying higher gene content, in the region. Both the general gene content and the potential medical interest make YACs from the high GC region a priority choice for long-range sequencing, and T bands on other chromosomes are presumptively attractive material for early sequence analysis. We have begun systematic genomic sequencing in the Xq28 region and have already confirmed, in agreement with the present results, that the GC content in the 30 kb around the G6PD gene is 57% (Chen *et al.*, 1991). The number of other genes and the GC content of the neighboring region have not yet been studied in detail, but two other genes, 3.5 (Toniolo *et al.*, 1988) and 4 kb long (Alcalay and Toniolo, 1988), have been reported thus far in the same zone; consistent with expectation, they are 61–62% GC.

One may wonder whether regional levels of GC content and gene-associated CpG islands are correlated by the action of a common mechanism during evolution. Aïssani and Bernardi (1991) have considered some speculative possibilities. It may be relevant that the very high GC levels and gene concentrations at a number of telomeres [or at intercalary T bands resulting from telomere fusion during evolution (Dutrillaux, 1973)] are paralleled by high transcriptional and recombinational activity (Bernardi, 1989, 1992; Rynditch *et al.*, 1991). Telomeres may localize at points along the nuclear membrane (Cremer *et al.*, 1982), and transcription and processing have been suggested to occur in organelle-like entities near nuclear pores (Newport and Forbes, 1987). Thus, by a mechanism that is still poorly defined, chromosomal activity that is generally promoted by localized regions of high GC and CpG island content might be augmented near telomeres.

ACKNOWLEDGMENTS

We thank Eric Green and our laboratory colleagues for discussions and comments on the manuscript. This study was supported by NIH Grant HG00201 and by grants from the Association Contre les Myopathies (AFM) and the French Ministry of Research and Technology. B.A. was an AFM Fellow in 1990–91.

REFERENCES

Aïssani, B., and Bernardi, G. (1991). CpG islands: Features and distribution in the genomes of vertebrates. *Gene* **106:** 173–183; 185–195.

Alcalay, M., and Toniolo, D. (1988). The mosaic genome of warm-blooded vertebrates. *Nucleic Acids Res.* **16:** 9527–9543.

Ambros, P. F., and Sumner, A. T. (1987). Correlation of pachytene chromomeres and metaphase bands of human chromosomes, and distinctive properties of telomeric regions. *Cytogenet. Cell Genet.* **44:** 223–228.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., *et al.* (1985). *Science* **228:** 953–958.

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

Bernardi, G. (1992). Genome organization and species formation in vertebrates. *J. Mol. Evol.*, in press.

Bettecken, T., Aissani, B., Muller, C. R., and Bernardi, G. (1992). Compositional mapping of the human dystrophin-encoding gene. *Gene* **122:** 329–335.

Bickmore, W. A., and Bird, A. P. (1992). Use of restriction endonucleases to detect and isolate genes from mammalian cells. *Methods Enzymol.* **216:** 224.

Bird, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3:** 342–347.

Brown, W. R. A., MacKinnon, P. J., Villasante, A., Spurr, N., Buckle, J., and Dobson, M. J. (1990). Structure and polymorphism of human telomere-associated DNA. *Cell* **63:** 119–132.

Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D., and Olson, M. V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244:** 1348–1351.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236:** 806–812.

Chen, E. Y., Cheng, A., Lee, A., Kuang, W. J., Hillier, L., Green, P., Schlessinger, D., Ciccodicola, A., and D'Urso, M. (1991). Sequence of human glucose-6-phosphate-dehydrogenase cloned in plasmids and a yeast artificial chromosome. *Genomics* **10:** 792–800.

Cremer, T., Cremer, C., Baumann, H., Luedtke, E.-K., Sperling, K., Teuber, V., *et al.* (1982). Rabl's model of the interphase chromosome arrangement tested in chinese hamster cells by premature chromosome condensation and laser-UV-Microbeam experiments. *Hum. Genet.* **60:** 46–56.

Davies, K. E., Mandel, J.-L., Monaco, A. P., Nussbaum, R. L., and Willard, H. F. (1991). Report on the genetic constitution of the X chromosome: Human Genome Mapping 11. *Cytogenet. Cell Genet.* **58:** 853–966.

DeSario, A., Aissani, B., and Bernardi, G. (1991). Compositional properties of telomeric regions from human chromosomes. *FEBS Lett.* **295:** 22–26.

Dietrich, A., Korn, B., and Poustka, A. (1992). *Mamm. Genome* **3:** 168–172.

Dutrillaux, B. (1973). Nouveau systeme de marquage chromosomique: Les bandes T. *Chromosoma* **41:** 395–402.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* **196:** 261–282.

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

Gorlin, J. B., Yamin, R., Egan, S., Stewart, M., Stossel, T. P., Kwuiatkowski, D. J., and Hartwig, J. H. (1990). Human endothelial actin-binding protein (ABP-280, nonmuscle filamin): A molecular leaf spring. *J. Cell Biol.* **111:** 1089–1105.

Huxley, C., Hagino, Y., Schlessinger, D., and Olson, M. V. (1991). The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion. *Genomics* **9:** 742–750.

Kaneko, K., Miyatake, T., Migeon, B. R., Warren, S. T., and Tsuji, S. (1992). Isolation of cosmid clones containing CpG islands at Xq24–qter region. *Cytogenet. Cell Genet.* **58:** 2069.

Krane, D. E., Hartl, D. L., and Ochman, H. (1991). Rapid determination of nucleotide content and its application to the study of genome structure. *Nucleic Acids Res.* **19:** 5181–5185.

Little, R. D., Pilia, G., Johnson, S., Zucchi, I., D'Urso, M., and Schlessinger, D. (1992). Yeast artificial chromosomes spanning 8 Mb and 10–15 centiMorgans of human cytogenetic band Xq26. *Proc. Natl. Acad. Sci. USA* **89:** 177–181.

Maestrini, E., Rivella, S., Tribioli, C., Purtilo, D., Rocchi, M., Archidiacono, N., *et al.* (1990). Probes for CpG islands on the distal long arm of the human X chromosome are clustered in Xq24 and Xq28. *Genomics* **8:** 664–670.

Montanaro, V., Casamassimi, A., D'Urso, M., Yoon, J.-Y., Freije, W., Schlessinger, D., Muenke, M., Nussbaum, R. L., Saccone, S., Maugeri, S., Santoro, A. M., Motta, S., and Della Valle, G. (1991). *In situ* hybridization to cytogenetic bands of yeast artificial chromosomes covering 50% of human Xq24–Xq28 DNA. *Am. J. Hum. Gen.* **48:** 183–194.

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

Newport, J. W., and Forbes, D. J. (1987). The nucleus: Structure, function, and dynamics. *Annu. Rev. Biochem.* **56:** 535–565.

Nguyen, C., Pontarotti, P., Birnbaum, D., Chimini, G., Rey, J. A., Mattei, J. F., *et al.* (1987). Large scale physical mapping in the q27 region of the human X chromosome: The coagulation factor IX gene and the mcf.2 transforming sequence are separated by at most 270 kilobase pairs and are surrounded by several 'HTF islands'. *EMBO J.* **6:** 3285–3289.

Nussbaum, R. L., Airhart, S. D., and Ledbetter, D. H. (1986). A rodent–human hybrid containing Xq24–Xqter translocated to a hamster chromosome expresses the Xq27 folate-sensitive fragile site. *Am. J. Med. Genet.* **23:** 457–466.

Palmieri, G., Cappa, V., Romano, G., D'Urso, M., Johnson, S., Schlessinger, D., Morris, P., Hopwood, J., Di Natale, P., Gatti, R., and Ballabio, A. (1992). The iduronate sulfatase gene: Isolation of a 1.2-Mb YAC contig spanning the entire gene and identification of heterogeneous deletions in patients with Hunter Syndrome. *Genomics* **12:** 52–57.

Palmieri, G., Romano, G., Casamassimi, A., D'Urso, M., Little, R. D., Abidi, F. E., Schlessinger, D., Lagerstrom, M., Malmgren, H., Steen-Bondeson, M. L., Pettersson, U., and Landegren, U. (1993). 1.5-Mb YAC contig in Xq28 formatted with sequence-tagged sites and including a region unstable in the clones. *Genomics* **16:** 586–592.

Poustka, A., Dietrich, A., Langestein, G., Toniolo, D., Warren, S., and Lehrach, H. (1991). *Proc. Natl. Acad. Sci. USA* **88:** 8302–8306.

Rynditch, A., Kadi, F., Geryk, J., Zoubak, S., Svoboda, J., and Bernardi, G. (1991). The isopycnic, compartmentalized integration of Rous sarcoma virus sequences. *Gene* **106:** 165–172.

Saccone, S., DeSario, A., Della Valle, G., and 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.

Schlessinger, D., Little, R. D., Freije, D., Abidi, F., Zucchi, I., Porta, G., Pilia, G., Nagaraja, R., Johnson, S. K., Yoon, J. Y., Srivastava, A., Kere, J., Palmieri, G., Ciccodicola, A., Montanaro, V., Romano, G., Casamassimi, A., and D'Urso, M. (1991). Yeast artificial chromosome-based genome mapping: Some lessons from Xq24–q28. *Genomics* **11:** 783–793.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., *et al.* (1992). The *C. elegans* genome sequencing project: A beginning. *Nature* **356:** 37–41.

Toniolo, D., Persico, M., and Alcalay, M. (1988). A "housekeeping" gene on the X chromosome encodes a protein similar to ubiquitin. *Proc. Natl. Acad. Sci. USA* **85:** 851–855.

Warren, S. T., Knight, S. J., Peters, J. F., Stayton, C. L., Consalez, G. G., and Zhang, F. P. (1990). Isolation of the human chromosomal band Xq28 within somatic cell hybrids by fragile X site breakage. *Proc. Natl. Acad. Sci. USA* **87:** 3856–3860.

Wolf, S. F., and Migeon, B. R. (1985). Clusters of CpG dinucleotides implicated by nuclease hypersensitivity as control elements of housekeeping genes. *Nature* **314:** 467–469.

Yunis, J. J. (1981). Mid-prophase human chromosomes. The attainment of 2000 bands. *Hum. Genet.* **56:** 293–298.

Zerial, M., Salinas, J., Filipski, J., and Bernardi, G. (1986). Gene distribution and nucleotide sequence organization in the human genome. *Eur. J. Biochem.* **160:** 479–485.