Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4884–4888, August 1984
Genetics

Characterization of a cloned DNA sequence that is present at centromeres of all human autosomes and the X chromosome and shows polymorphic variation

(C-band heterochromatin/highly repetitive DNA/segregation analysis/chromosome mapping/amplified sequences)

ETHYLIN WANG JABS, STANLEY F. WOLF, AND BARBARA R. MIGEON

Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

Communicated by John W. Littlefield, April 18, 1984

ABSTRACT **We have identified a human DNA recombinant (p308) with a 3.0-kilobase (kb) *Bam*HI insert that hybridizes *in situ* exclusively to the centromeric region of all human autosomes and the X chromosome. This highly repetitive sequence is significantly enriched on several chromosomes, most prominently on chromosome 6. In all individuals, the majority of genomic repeats are organized as tandem 3.0-kb *Bam*HI repeats, each containing one *Taq* I site; the others are organized into *Bam*HI and *Taq* I repeats of variable size that have some chromosome specificity. Using mouse–human hybrids, we have defined the specific organization of this sequence on chromosomes 6, 3, and X. In some individuals, there are differences in the number and nature of the tandem repeats. These polymorphisms segregate in families as if chromosome specific. Although variable from one chromosome to another, 308 contains sequences homologous to DNA present in centric heterochromatin of essentially all human chromosomes and is evolutionarily conserved. Therefore, a significant component of pericentric DNA is similar for all human chromosomes.**

The centromere of mammalian chromosomes plays an important role in cell division, as proper function is required for chromosomes to be stably transmitted from one cell to its progeny. Sequences with centromeric function in lower eukaryotes are composed of single copy as well as low repetitive sequences (1). However, the region around the centromere of mammalian chromosomes consists of highly repetitive DNA. The role of this abundant material is unknown, but it has been shown to be heterogeneous (2–4) and to exhibit staining properties characteristic of heterochromatin (5).

Our knowledge of the organization of human centromeric heterochromatin is mainly derived from studies of satellite DNAs. Gosden *et al.* (6, 7), Jones (8), and Manuelidis (9, 10) showed that each of the satellite DNAs obtained from density gradients localizes to a different set of human centromeres, but none is exclusively confined to centromeres. Based on restriction endonuclease digestion, satellite DNA has been shown to consist of a variety of repeated sequences (11, 12). Because satellite DNA is heterogeneous, we know little about the organization of pericentric DNA on individual chromosomes. Only a few cloned sequences from the centromeric regions of human chromosomes have been studied. Cooke and Hindley (13) cloned an *Eco*RI fragment from human satellite DNA that hybridizes mainly to the C-band region of chromosome 1 but has some homology to C bands of other chromosomes (14). Yang *et al.* (15) identified a 2.0-kilobase (kb) *Bam*HI fragment that is homologous to sequences in human satellite III DNA and to the alphoid repeat DNA family (16). The major site of hybridization is to the centromere of the X chromosome, but there is homology to DNA of other human centromeres as well (15, 17).

We have identified another highly repetitive human DNA sequence that localizes predominantly, if not exclusively, to centromeres. We have shown that individual chromosomes differ with respect to the number and size of repeats and that these differences provide a marker for heterochromatin on the X chromosome and autosomes. Our studies have revealed the presence of additional repeats of different sizes in many individuals, and these polymorphic variants should be useful for linkage studies.

MATERIALS AND METHODS

Source of Mammalian DNA. DNA was purified from human placentas, lymphoblasts, lymphocytes, and hybrid fibroblasts as described by Wolf *et al.* (18). Lymphoblast cultures with 46 X, isodicentric X (19), and 46 XY, 1 qh^+ (20) were established with Epstein–Barr virus in our laboratory. Mouse DNAs were obtained from A9 cells (21), from SVtfm fibroblasts (GM5267), recently established mouse cells of renal origin (22), and from mouse C57 BL/6J leukocytes. The mouse–human hybrid clone 91b with human X centromere (X^+ hybrid) and 6TG-resistant derivatives (X^- hybrid) have been described (18) and the hybrid clones with centromeres of chromosomes 3 or 6 were derived in this laboratory from a variety of independent fusions between mouse A9 fibroblasts and human parental fibroblasts from different individuals.

Source of Recombinant Plasmid. p308, a 3.0-kb *Bam*HI human DNA fragment cloned into pBR322, was obtained from a human recombinant library (18, 23). The 308 insert was purified by the method of Vogelstein and Gillespie (24).

Southern Blot Analysis. DNA (5 μg per lane) was digested with restriction enzymes (16 hr; as suggested by the manufacturers, Bethesda Research Laboratories and New England Biolabs), electrophoresed, transferred, and hybridized with ^{32}P-labeled nick-translated DNA (10^7 cpm/μg; 5–50 $\times$ 10^{-3} μg/ml) (18). All filters were washed under stringent conditions in 0.1$\times$ SET at 68°C, unless otherwise noted. Autoradiographs were obtained using Kodak X-Omat R film and Cronex Lightning Plus intensifying screens, following exposure times ranging from 15 min for human genomic DNA to several weeks for hybrids.

***In Situ* Hybridization.** ^{3}H-labeled p308 DNA (10^7 cpm/μg; 0.1 μg/ml) was hybridized *in situ* (25) to metaphase chromosomes obtained from human lymphocytes of normal males and females and human lymphoblasts of 46 X, isodicentric X, and 46 XY, 1 qh^+ karyotypes. Chromosomes were identified prior to hybridization by quinacrine banding (26).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobase(s).

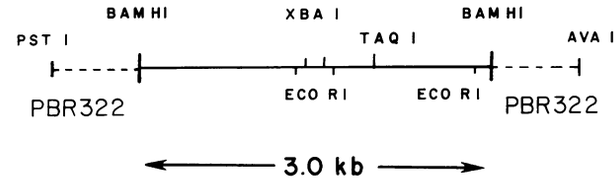

FIG. 1. Restriction enzyme map of p308. There are no *Ava* I, *Hin*dIII, *Kpn* I, *Pst* I, or *Sst* I sites within the insert.

RESULTS

Characterization of 308. A map of p308, a pBR322 recombinant with a 3.0-kb *Bam*HI human DNA repetitive sequence (23), is shown in Fig. 1. When the insert was cut with restriction enzymes shown in Fig. 1, each resultant fragment hybridized strongly to the ^{32}P-labeled 46 XY, genomic probe as if it was repetitive. We estimate that there are $\approx 10^5$ copies of this sequence in the diploid genome based on DNA titration and relative hybridization of a single copy probe, 307 (18). No cross hybridization with sequences of the Alu family was observed, even under decreased stringency (23). The temperature of denaturation [determined by the method of Roshash *et al.* (27)] is 50–55°C for 308 compared to 60–65°C for pBR322. The G+C content of pBR322 is 54%, so therefore, 308 is relatively A+T rich.

Localization of 308 by *in Situ* Hybridization. Sequences homologous to 308 are present in centromeric regions of autosomes and the X chromosome (Fig. 2), but the extent of hybridization varies among chromosomes. Fig. 3 shows the relative intensity of hybridization for each chromosome in 12-day autoradiographs. About 40% of the grains hybridize to chromosome 6, while chromosomes 1, 3, 12, and 19 are significantly more labeled than others. Longer exposures confirm that the centromeres of all autosomes and the X chromosome are labeled. However, there is no hybridization to the Y chromosome under conditions where all other chromosomes are labeled. These results do not exclude the possibility of significantly fewer copies of this sequence on the Y chromosome.

Hybridization of 308 was not proportional to the length of C-band regions, as there is no increased hybridization to the large C band of chromosome 9 (Fig. 3) and to the amplified C

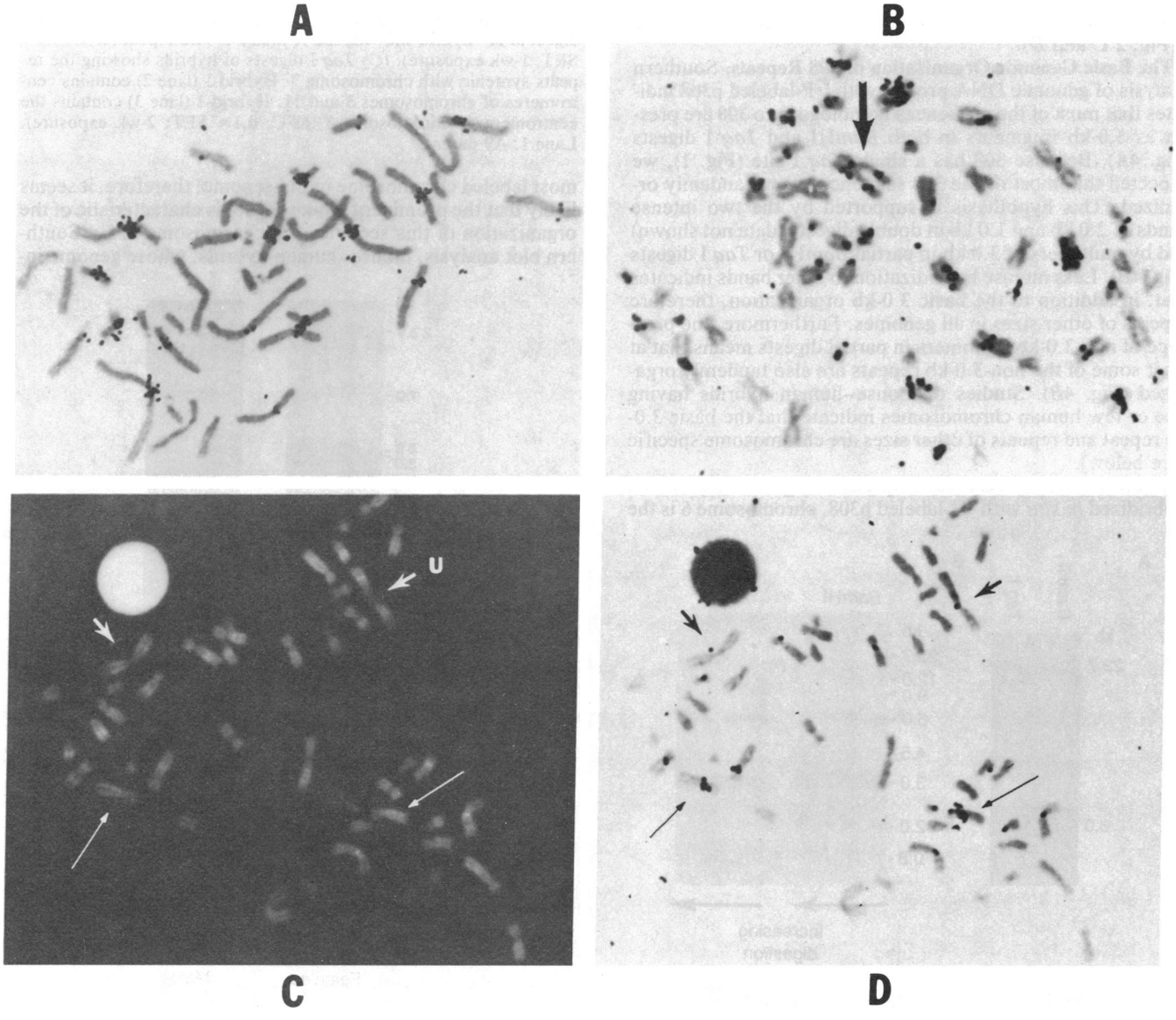

FIG. 2. *In situ* hybridization of ^{3}H-labeled p308 to human centromeres. (*A*) 46 XY; (*B*) 46 X, isodicentric X. Note the isodicentric X chromosome (arrow) is labeled at both centromeres. (*C* and *D*) 46 XY, 1 qh$^+$ *uncoiler locus* stained with quinacrine (*C*) and hybridized (*D*). Chromosomes 6 (long arrows) are most labeled while chromosomes 1 (short arrows) are labeled, but the *uncoiler locus* (u) is not.

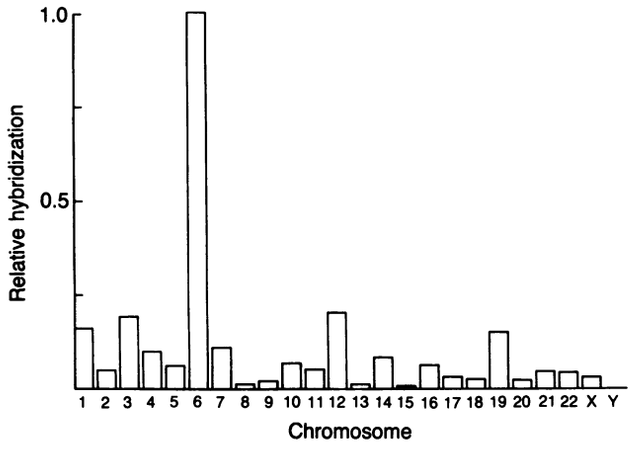

FIG. 3. Relative *in situ* hybridization of ^{3}H-labeled p308 to specific human chromosomes. Relative hybridization is the mean number of grains at the centromere of each chromosome relative to that at the centromere of chromosome 6, from analysis of 25 metaphases (2 males and 2 females) in 12-day autoradiographs.

band associated with the *uncoiler locus* (20) on chromosome 1 (Fig. 2 *C* and *D*).

The Basic Genomic Organization of 308 Repeats. Southern analysis of genomic DNA probed with ^{32}P-labeled p308 indicates that most of the sequences homologous to 308 are present as 3.0-kb fragments in both *Bam*HI and *Taq* I digests (Fig. 4*A*). Because 308 has a single *Taq* I site (Fig. 1), we expected that most of the 308 sequences were tandemly organized. This hypothesis is supported by the two intense bands of 2.0 kb and 1.0 kb in double digests (data not shown) and by multimers of 3.0 kb in partial *Bam*HI or *Taq* I digests (Fig. 4*B*). Less intense hybridization to other bands indicates that, in addition to the basic 3.0-kb organization, there are repeats of other sizes in all genomes. Furthermore, the presence of non-3.0-kb multimers in partial digests means that at least some of the non-3.0-kb repeats are also tandemly organized (Fig. 4*B*). Studies of mouse–human hybrids having one or few human chromosomes indicate that the basic 3.0-kb repeat and repeats of other sizes are chromosome specific (see below).

Evidence for Chromosome-Specific Organization. When hybridized *in situ* with ^{3}H-labeled p308, chromosome 6 is the most labeled chromosome in the genome; therefore, it seems likely that the prominent 3.0-kb repeat is characteristic of the organization of this sequence on chromosome 6. By Southern blot analysis, mouse–human hybrids, whose genome in-

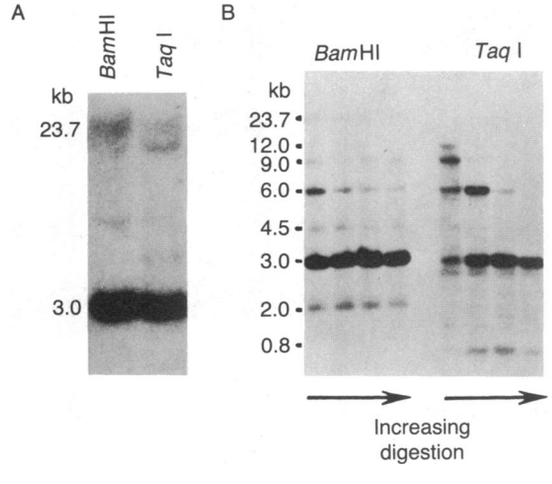

FIG. 4. 308 is mainly organized as 3.0-kb *Bam*HI or *Taq* I tandem repeats. 46 XY DNA (5 μg) was digested with *Bam*HI or *Taq* I, blotted, probed with ^{32}P-labeled p308 (0.005 μg/ml) washed at 68°C in 0.1× SET, and autoradiographed for 15 min. (*A*) Complete digests; (*B*) partial digests, showing multimers of the 3.0-kb band.

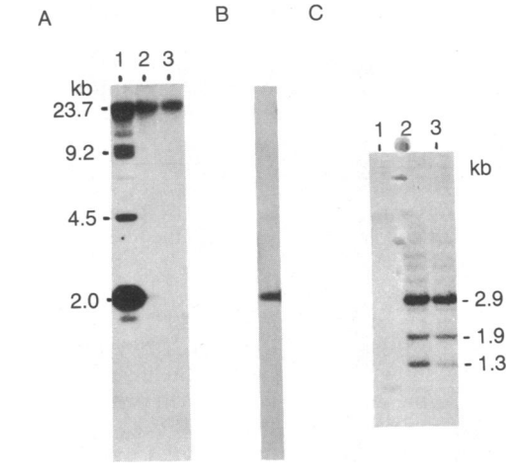

FIG. 5. Organization of 308 is chromosome specific. Southern blots of DNA from mouse–human hybrids probed with ^{32}P-labeled p308 (0.05 μg/ml). (*A*) *Bam*HI digests of X^+ hybrid (lane 1) with only the human X chromosome, derivative X^- hybrid (lane 2), and A9 parent (lane 3), showing the 2.0-, 4.5-, and 9.2-kb bands that are on the X chromosome (55°C; 1× SET; 3-day exposure). (*B*) *Bam*HI digest of X^+ hybrid showing the X-linked 2.0-kb repeat (68°C; 0.1× SET; 1-wk exposure). (*C*) *Taq* I digests of hybrids showing the repeats syntenic with chromosome 3. Hybrid 1 (lane 2) contains centromeres of chromosomes 3 and 11. Hybrid 2 (lane 3) contains the centromere of chromosome 3 (68°C; 0.1× SET; 2-wk exposure). Lane 1, A9 parent.

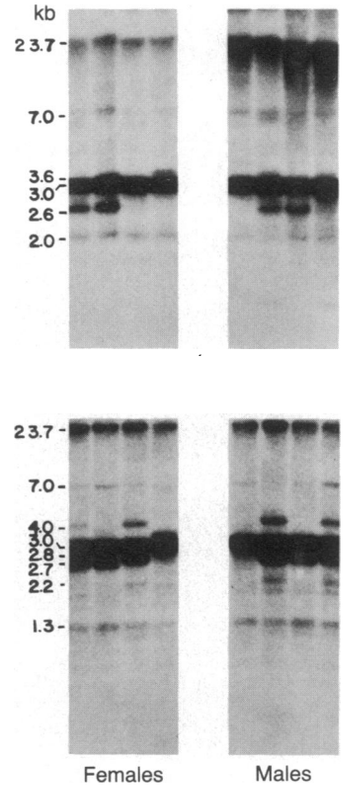

FIG. 6. *Bam*HI and *Taq* I polymorphisms. Autoradiographs of genomic DNAs from unrelated females and males (4 each), digested with *Bam*HI (*Upper*) and *Taq* I (*Lower*), probed with ^{32}P-labeled p308.

cludes chromosome 6, have the 3.0-kb repeat, which is absent in hybrids lacking chromosome 6.

Because 308 hybridizes predominantly to chromosome 6, it is difficult to determine the organization of this sequence on other chromosomes from studies of human genomic DNA. Mouse–human hybrids with only a few human chromosomes provide the means to examine directly the organization of this sequence on chromosomes other than chromosome 6. Studies of hybrids with the X as the only human chromosome present indicate that the 2.0-kb *Bam*HI repeat is X-linked (Fig. 5 *A* and *B*), and hybrids with more than one X show dosage effect. Furthermore, *Taq* I bands of 2.9, 1.9, and 1.3 kb are seen in hybrids with centromeres of human chromosome 3 (Fig. 5*C*) but are not present in hybrids lacking the chromosome 3 centromere.

Some Repeats Are Polymorphic. In addition to the basic repeats described above that are found in all individuals, some individuals have other prominent *Bam*HI and *Taq* I repeats (Fig. 6). These extra bands represent qualitative and quantitative polymorphic variants. To determine the extent of this variability, we analyzed DNA from unrelated individuals (Fig. 7 *A* and *B*). Prominent *Bam*HI variants of 7.0, 3.6, and 2.6 kb and *Taq* I variants of 4.0, 3.5, 2.8, 2.6, and 2.5 kb were observed. Many individuals had more than a single variant, and the patterns along with their frequencies are shown in Fig. 7. The cosegregation of two variants (i.e., the 2.6-kb *Bam*HI and 2.6-kb *Taq* I bands) suggests that they are probably syntenic and may, in fact, reflect the same repeat. From these patterns, we estimated the frequency of specific polymorphic bands. These ranged from 5% for the 2.6-kb *Taq* I fragment(s) to 67% for the 2.8-kb *Bam*HI fragment(s).

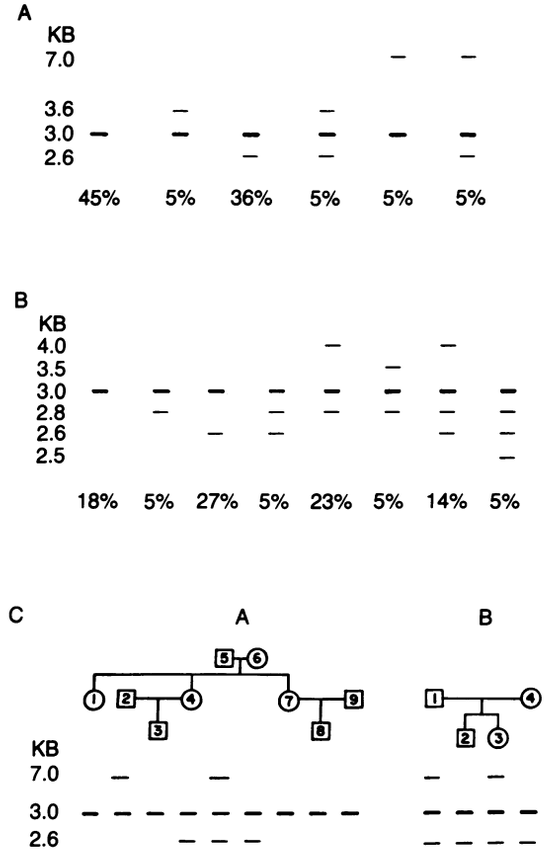

FIG. 7. Summary of prominent polymorphic band patterns. Representation of blots of genomic DNA probed with p308 and frequency of pattern (%). DNA from 22 unrelated individuals (5 Blacks, 16 Caucasians, 1 Oriental) was digested with *Bam*HI (*A*) and *Taq* I (*B*). (*C*) Pedigrees showing segregation of *Bam*HI polymorphisms.

The intensity of a polymorphic band may vary among individuals, perhaps reflecting variation in the number of repeats and/or heterogeneous origin.

Polymorphisms Segregate in Families. To test the possibility that each specific polymorphic variant is of single chromosome origin, we studied segregation of polymorphic variants in four families. Our observations of the two informative pedigrees (Fig. 7*C*) are consistent with the hypothesis for the following reasons. If the parents have a variant, the child does not necessarily inherit it (pedigree A; individuals 2–4). If neither parent has the variant band, the child does not have the variant (pedigree A; individuals 7–9). If a child has a variant, then at least one of his parents has the variant (pedigree B; individuals 1 and 3), and multiple variant bands may be transmitted independently (pedigree B; individuals 1–4). Furthermore, based on the intensity of the variant bands the dosage is as expected if the individual received the variant from one or both parents. Therefore, the polymorphic bands segregate as though each variant is present as tandem repeats, on one or both members of a chromosome pair.

Evolutionary Conservation of 308. There are sequences homologous to 308 in the genomes of mouse leukocytes and fibroblasts, but hybridization is much less intense indicating that these sequences are either less reiterated or significantly less homologous. As a consequence, using conditions of *in situ* hybridization that detect repetitive sequences on human chromosomes, we observed no labeled mouse chromosome. However, some 308 repeats in the murine genome are present in a 3.0-kb *Bam*HI and *Taq* I organization (Fig. 8).

DISCUSSION

Previous studies of cloned pericentric DNA sequences (13, 15, 17) have emphasized their chromosome specificity so that centric heterochromatin is perceived to be highly heterogeneous. The results of our studies, on the other hand,

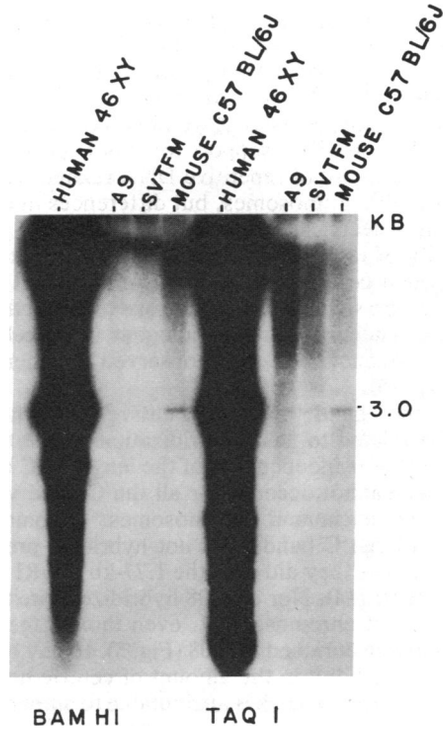

FIG. 8. 308 is evolutionarily conserved. Southern blots from human leukocytes (46 XY), mouse fibroblasts (A9 and SVtfm), and murine leukocytes (C57 BL/67) digested with *Bam*HI and *Taq* I, probed with ^{32}P-labeled p308, washed at 68°C in 0.1× SET, and exposed for 2 weeks.

suggest that a significant component of the repetitive DNA around the centromere is similar for all human chromosomes.

We have shown that the tandemly repeated 308 sequence is present at the centromere of all autosomes and the human X chromosome. From the segregation of restriction fragment patterns and variations in the intensity of hybridization of the sequence to chromosome-specific DNA, it is clear that there are significant differences in the organization of the sequence on various chromosomes, indicating some divergence has occurred. Differences of this kind have previously obscured the relationship among these pericentric sequences.

Our studies indicate that there is a family of tandem repeats that are prominent in centric heterochromatin of all chromosomes. Although predominant on chromosome 6, the 308 sequence is also prominent on chromosomes 1, 3, 12, and 19. Furthermore, it is well represented on the rest of the chromosomes because it is observed by *in situ* hybridization under conditions where only reiterated sequences can be detected. Those chromosomes that are less intensely labeled *in situ* may have comparable numbers of copies of related sequences. For example, the X chromosome does not hybridize with the intensity of chromosome 6 *in situ* (Fig. 3) or in Southern blots. However, it is likely that 308 is related to the X-enriched 2.0-kb *Bam*HI repeat described by Yang *et al.* (15) and by Willard *et al.* (17) for the following reasons. The X-enriched sequence hybridizes *in situ* to centromeres of at least some autosomes (15) and to a 3.0-kb *Bam*HI band in Southern blots of genomic DNA. Furthermore, like the X-enriched probe (15), 308 hybridizes to 2.0-kb *Pst* I (data not shown) and *Bam*HI tandem repeats on the X chromosome. Presumably, the cloned 2.0-kb *Bam*HI sequence is derived from X-centric heterochromatin, while 308 comes from chromosome 6. The signal for hybridization *in situ* of the cloned 2.0-kb *Bam*HI fragment to the X chromosome (15) is similar to that observed for 308 to chromosome 6, suggesting that the number of related sequences is similar on both chromosomes. The decreased hybridization of 308 to the X chromosome most likely reflects divergence of the sequences on chromosome 6 and on the X chromosome. The increased hybridization to X sequences that we observed at lower stringency (Fig. 5*A*) provides support for this hypothesis. Furthermore, the amount of repetitive DNA related to 308 may be similar for all chromosomes, but differences in degree of homology may decrease the hybridization signal.

This family of tandem repeats also may be related to sequences around centromeres of other mammalian chromosomes. There are sequences homologous to 308 in the mouse genome, and tandem repeats homologous to the related 2.0-kb *Bam*HI sequence have been observed in other primate genomes (15, 16).

On the other hand, not all repetitive pericentric DNA seems to be related to 308. Hybridization of 308 to human pericentric DNA is independent of the length of C bands, so this sequence cannot account for all the C-band variability that characterizes human chromosomes. Chromosomes 9 and 16, with large C bands, did not hybridize prominently with 308, whereas they did with the 1.77-kb *Eco*RI sequence from satellite III (14). Nor did 308 hybridize more to the enlarged C band of chromosome 1, even though the chromosome is relatively enriched for 308 (Fig. 3). It may be that the cytogenetic variability in the amount of centric heterochromatin among chromosomes is attributable to amplification of a subset of the sequences present in the pericentric DNA.

Our studies have revealed that there is significant polymorphic variation in the organization of DNA homologous to 308 among normal individuals. These polymorphisms segregate as if chromosome specific and may account for at least some of the C-band heteromorphisms. On the other hand, we have not observed an association between cytogenetic markers and the presence of the variants of 308. It is not yet clear whether the prominent polymorphic bands represent a significant increase in the total amount of homologous sequences in the genome or a change in distribution of closely related sequences among various bands. The intensity of the bands and distribution of grains *in situ* suggest that the polymorphism may actually be from chromosome 6. It would be of interest to learn if the related sequences on the X and other chromosomes also show polymorphic variation. Such polymorphic sequences should provide useful centromeric markers, especially for chromosomes without C-band heteromorphisms. Because 308 contains sequences that are so ubiquitous and evolutionarily conserved, it may be functional and should provide the means of identifying chromatin components important for centromeric function.

Note Added in Proof. From recent studies of hybrids, we know that the Y chromosome has sequences homologous to 308.

We gratefully acknowledge the assistance of Pamela J. Hubberman for *in situ* hybridization studies and Lori A. Garrison in preparation of the manuscript. These studies were supported by National Institutes of Health Grant HD 05465. While initiating these studies, E.W.J. was a Stetler Research Fellow. She is currently supported by a National Institutes of Health postdoctoral fellowship.

1. Fitzgerald-Hayes, M., Clarke, L. & Carbon, J. (1982) *Cell* **29,** 235–244.
2. John, B. & Miklos, G. L. (1979) *Int. Rev. Cytol.* **58,** 1–114.
3. Brutlag, D. L. (1980) *Annu. Rev. Genet.* **14,** 121–144.
4. Singer, M. F. (1982) *Int. Rev. Cytol.* **76,** 67–193.
5. Comings, D. E. (1978) *Annu. Rev. Genet.* **12,** 25–46.
6. Gosden, J. R., Mitchell, A. R., Buckland, R. A., Clayton, R. P. & Evans, H. J. (1975) *Exp. Cell Res.* **92,** 148–158.
7. Gosden, J. R., Mitchell, A. R., Seuanez, H. N. & Gosden, C. M. (1977) *Chromosoma* **63,** 253–271.
8. Jones, K. W. (1977) in *Molecular Structure of Human Chromosomes*, ed. Yunis, J. (Academic, New York), pp. 295–326.
9. Manuelidis, L. (1978) *Chromosoma* **66,** 1–21.
10. Manuelidis, L. (1978) *Chromosoma* **66,** 23–32.
11. Mitchell, A. R., Beauchamp, R. S. & Bostock, C. J. (1979) *J. Mol. Biol.* **135,** 127–149.
12. Frommer, M., Prosser, J., Tkachuk, D., Reisner, A. H. & Vincent, P. C. (1982) *Nucleic Acids Res.* **10,** 547–563.
13. Cooke, H. J. & Hindley, J. (1979) *Nucleic Acids Res.* **6,** 3177–3197.
14. Gosden, J. R., Lawrie, S. S. & Cooke, H. J. (1981) *Cytogenet. Cell Genet.* **29,** 32–39.
15. Yang, T. P., Hansen, S. K., Oishi, K. K., Ryder, O. A. & Hamkalo, B. A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6593–6597.
16. Maio, J. J., Brown, F. L. & Musich, P. R. (1981) *Chromosoma* **83,** 103–125.
17. Willard, H. F., Smith, K. D. & Sutherland, J. (1983) *Nucleic Acids Res.* **11,** 2017–2033.
18. Wolf, S. F., Mareni, C. E. & Migeon, B. R. (1980) *Cell* **21,** 95–102.
19. Sarto, G. E. & Therman, E. (1980) *Am. J. Obstet. Gynecol.* **136,** 904–911.
20. Donahue, R. P., Bias, W. B., Renwick, J. H. & McKusick, V. A. (1968) *Proc. Natl. Acad. Sci. USA* **61,** 949–955.
21. Littlefield, J. W. (1964) *Science* **145,** 709.
22. Migeon, B. R., Brown, T. R., Axelman, J. & Migeon, C. J (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6339–6343.
23. Jabs, E. W., Wolf, S. F. & Migeon, B. R. (1984) *Somatic Cell Genet. Mol. Biol.* **10,** 93–103.
24. Vogelstein, B. & Gillespie, D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 615–619.
25. Harper, M. E., Ullrich, A. & Saunders, G. F. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 4458–4460.
26. Caspersson, T., Lomakka, G. & Zech, L. (1971) *Hereditas* **67,** 89–102.
27. Roshash, M., Blank, D., Fahrner, K., Hereford, L., Ricciardi, R., Roberts, B., Ruby, S. & Woolfard, J. (1979) *Methods Enzymol.* **68,** 454–469.