Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 1080–1084, February 1992
Genetics

Distribution of interspersed repeats (*Ålu* and *Kpn*) on *Not* I restriction fragments of human chromosome 21

(pulsed-field gel electrophoresis/physical mapping/genomic organization)

JESUS SAINZ*, LARYSA PEVNY†, YUE WU‡, CHARLES R. CANTOR*‡, AND CASSANDRA L. SMITH*‡§

*Division of Chemical Biodynamics, Lawrence Berkeley Laboratory and ‡Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720; and †Department of Genetics and Development, College of Physicians and Surgeons, Columbia University, New York, NY 10032

Contributed by Charles R. Cantor, November 7, 1991

ABSTRACT **Interspersed repeated sequences (*Alu* and *Kpn*) were used as probes to detect a set of *Not* I restriction fragments of human chromosome 21 from the hybrid cell line WAV17. Forty different *Not* I fragments, ranging in size from <0.05 megabase (Mb) to 7.0 Mb, were identified. The total length of these fragments was 47.3 Mb. This length provides an estimate of the minimum size of the chromosome and a minimum number of fragments to be ordered to create a complete restriction map. The average length *Not* I fragment is 1.2 Mb. *Alu* and *Kpn* fragments are not always coincident: a 2.9-Mb fragment is detected with *Kpn* but not with *Alu*, and 13 fragments, ranging from <0.05 Mb to 5.6 Mb, are detected with *Alu* but not with *Kpn*; the 26 remaining fragments, covering 75% (35.3 Mb) of the total length, are detected with both repetitive probes. The presence of so many noncoincident fragments and the high variation of the hybridization signal intensities of the fragments suggest a very nonuniform distribution of *Kpn* and *Alu* repeats.**

Human chromosome 21, the smallest human chromosome, is an appropriate model for large-scale chromosome mapping. This acrocentric chromosome contains rRNA-encoding DNA repeats and contains genes responsible for Down syndrome (1–3), some forms of familial Alzheimer disease (4–6), and amyotrophic lateral sclerosis (7).

Previous restriction mapping efforts on smaller chromosomes were aided considerably by the *a priori* knowledge of all the DNA fragments that needed to be ordered. The restriction enzyme *Not* I cleaves the human genome into fragments with an estimated average size of 1.0 megabase (Mb) (8, 9). Since human chromosome 21 is estimated to be ≈51 Mb (10, 11), this chromosome should contain roughly 51 *Not* I fragments. Chromosome 21 *Not* I restriction fragments, fractionated by pulsed-field gel electrophoresis (PFGE; refs. 12 and 13), can be identified in human–rodent hybrid cell line DNA by using human-specific interspersed repeated sequences as hybridization probes (9).

Short and long interspersed repeated DNA sequences (SINEs and LINEs, respectively) are classified by their length (14, 15). Some studies suggest that there is a complementary distribution of the SINEs and LINEs in the human genome (16–18). In human DNA, the major SINE is the *Alu* sequence family (19). The human genome contains 300,000 to 900,000 copies of the 300-base-pair (bp) *Alu* sequence (20–22). The *Alu* family represents 3–9% of the haploid genome (21). The average distance between *Alu* sequences is 3–10 kilobases (kb), and copies appear to be widely distributed across the entire human genome (14, 23, 24).

The major human LINE family is the 6.4-kb L1 or *Kpn* sequence. Many *Kpn* repeats have deletions and rearrangements, especially at the 5′ end, but most of them share a common 3′ end. The number of copies is 50,000–100,000 for the 3′ end and 4000–20,000 for the 5′ end (22, 25), with an average distance between elements of 30–60 and 150–750 kb, respectively. It is not clear whether the LINE sequences containing and lacking the 5′-end regions are distributed randomly (17).

From estimates of interspersed repeat copy numbers we can calculate that, on average, each megabase of human DNA should contain 100–300 copies of *Alu*, 17–32 copies of the 3′-end *Kpn* sequence, and 1.3–6.7 copies of the 5′-end *Kpn* sequence. Thus, the use of *Alu* repeats as hybridization probes should detect, with high probability, all the chromosome 21 *Not* I restriction fragments. For instance, Poisson statistics predict that the probability of not finding an *Alu* sequence on a 1.0-Mb fragment is $<4 \times 10^{-44}$. The probability of not finding a 3′-end *Kpn* sequence on a 1.0-Mb fragment is $<6 \times 10^{-8}$, and the probability of not finding a 5′-end *Kpn* sequence will be somewhere in the range of 0.001–0.26. Thus, when used in hybridization these probes should, in concert, detect virtually all the *Not* I fragments of human chromosome 21, unless there are peculiar regions where some repeats are not tolerated.

MATERIALS AND METHODS

Hybrid cell line WAV17 is a mouse cell line that contains human chromosome 21 as its only human component. WAV17 has an average of three copies of human chromosome 21 per cell (ref. 26; data not shown). DNA from this cell line was purified in agarose blocks (inserts), digested with the restriction enzyme *Not* I (10 units of enzyme per μg of DNA), and fractionated by PFGE (27, 28). Size standards were bacteriophage λ DNA concatemers, and *Saccharomyces cerevisiae, Pichia*, and *Schizosaccharomyces pombe* chromosomal DNAs (refs. 29 and 30; T.-Y. Zhang, S. Ringquist, J.-B. Fan, and C.L.S., unpublished data). Southern blotting and hybridization conditions were as described (27, 28). Two different *Alu* probes were used. Plasmid pX-H contains a 2.0-kb *Xba* I/*Hin*dIII fragment that has five different *Alu* sequences (31). Plasmid Blur8 has 300 bp of DNA containing a single *Alu* sequence (32). The *Alu* sequences were analyzed by using an *Alu* identification program.¶ This program compares *Alu* sequences with an *Alu* consensus sequence and six different *Alu* subfamilies (34). Three *Kpn* probes were used (see Fig. 1 and ref. 35). Tεγ-C (1.6 kb) contains 1.2 kb of 3′-end L1 sequence, Tεγ-D (1.5 kb) contains 0.8 kb of sequence adjacent to Tεγ-C, and ψ^{LTR} (0.45 kb) contains sequences near the 5′ end of an L1 element including features

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.

Abbreviations: Mb, megabase(s); PFGE, pulsed-field gel electrophoresis; LINE, long interspersed repeated DNA sequence; SINE, short interspersed repeated DNA sequence.
§To whom reprint requests should be addressed.
¶Milosavjlevic, A., Jurka, J. & Monardo, P. *Human Alu Sequence Identification Via Electronic Mail* (Internet: pythia@cis.ucsc.edu).

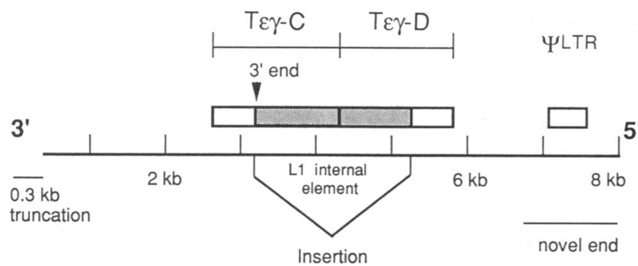

FIG. 1. Location of the *Kpn* probes (Tεγ-C, Tεγ-D, ψ^{LTR}) on an L1 element. Shaded portion indicates an inserted L1 element lacking the 5′ end. The figure is a redrawn composite from Rogan *et al.* (35).

that resemble the long terminal repeat of transposable DNA elements. Probes were labeled with [^{32}P]dCTP by random priming (36). Band signal intensities were quantitated directly from the labeled membrane by using a Molecular Dynamics 400A PhosphorImager image scanner. Differences in signal intensity for the same probe hybridized to different membranes are seen due to experimental factors such as the amount of DNA loaded, time of exposure, quality of the blot, and specific activity of the probe. These differences were compensated for by integrating the total amount of signal on each membrane with a particular probe before calculating the relative amount of signal assigned to individual bands.

RESULTS AND DISCUSSION

Hybridization of a PFGE-fractionated *Not* I digest of WAV17 DNA, which contains chromosome 21 as its only human component, with the *Alu* probe Blur8 (containing a single *Alu* sequence) identified a discrete set of bands (Fig. 2). Computational analysis (of the Blur8 *Alu* sequence) using an *Alu* identification program¶ revealed that it is 91% homologous to the *Alu* consensus sequence and belongs to the largest *Alu* subfamily, Sx (34). Hybridization experiments using *Alu* probe pX-H (containing five *Alu* sequences) detected almost the same number of bands as probe Blur8. However, two bands (0.20 and 0.087 Mb) were detected only with probe pX-H (Table 1). The larger number of bands detected with pX-H is consistent with the fact that it contains five different *Alu* sequences (three, one, and one copy of subfamilies J, Sp, and Sc, respectively). The pX-H *Alu* sequences have 68–84% homology with the *Alu* consensus sequence (34). The *Alu* probes detected a total of 39 *Not* I bands, ranging in size from 0.014 to 7.0 Mb, with a total apparent length of 44.4 Mb. The largest size standard used was the 5.7-Mb chromosome I of *Sc. pombe* (30); the size of the largest *Not* I fragment from chromosome 21 is >5.7 Mb. Digestion of this *Not* I fragment with other restriction enzymes revealed its true size of 7 Mb (R. Oliva, J.S., H. Ichikawa, M. Murata, M. Ohki, and C.L.S., unpublished data). The average size of the bands detected with pX-H is 1.1 Mb. Seventy-two percent of the bands detected with the *Alu* probe pX-H are <1 Mb.

As expected, the *Kpn* probes detected a smaller number of bands than the *Alu* probes. All the bands detected using the *Kpn* probes were coincident in size with the bands detected using *Alu* probes, with one exception—a 2.9-Mb band (Table 1). The 3′-end *Kpn* probe Tεγ-C detected 27 bands ranging in size from 0.14 to 7.0 Mb with a total apparent length of 38.2 Mb and an average size of 1.4 Mb (Fig. 2*B*). Almost all of the large *Not* I bands were detected with both *Alu* and *Kpn* probes. Of the 12 different bands >1.0 Mb, 11 bands were detected with *Alu* probes and 11 bands were detected with *Kpn* probes; 10 bands were detected with both probes. The *Kpn* probes hybridized to a 2.9-Mb fragment that was not detected with the *Alu* probes (Fig. 3), and the *Alu* probes detected a 5.6-Mb band not seen with the *Kpn* probes. However, there is less coincidence with LINE and SINE probes for *Not* I bands <1.0 Mb. For instance, while the *Alu* probes detected 28 such bands, only 16 of them were seen with the *Kpn* probes. These results are consistent with the lower copy number of *Kpn* sequences. The probability of missing a *Not* I fragment <1.0 Mb with a *Kpn* probe is clearly higher than the probability with an *Alu* probe. For example,

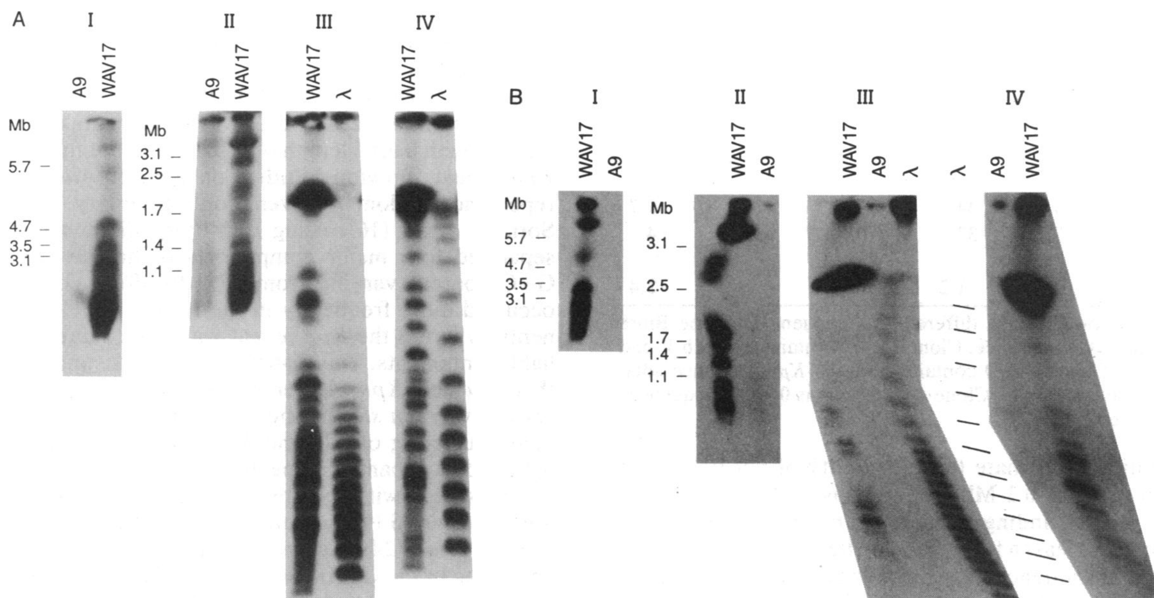

FIG. 2. *Not* I fragments detected with *Alu* and *Kpn* probes. (*A*) *Alu*-containing *Not* I bands detected by hybridization with Blur8. (*B*) *Kpn*-containing *Not* I bands detected by hybridization with Tεγ-C. *Not* I-digested DNA from mouse hybrid cell line WAV17 (containing human chromosome 21) and mouse cell line A9 was fractionated by PFGE and hybridized as described. Pulse times of 60 sec (*A*, lanes IV and *B*, lanes IV), 120 sec (*A*, lanes III and *B*, lanes III), 800 sec (*B*, lanes II), or 6600 sec (*B*, lanes I) and pulse time programs of 4500, 3600, 2700, and 1800 sec (*A*, lanes I), or 2700, 1800, and 900 sec (*A*, lanes II) were used. The field strengths were 10 and 9 V/cm at 11°C [90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3 (TBE buffer) temperature] (lanes III and IV), and 3.3 and 2.6 V/cm at 8°C (TBE buffer temperature) (lanes I and II). Separation ranges were 3.0–6.0 Mb (lanes I), 1.0–3.0 Mb (lanes II), 0.5–1.0 Mb (lanes III), and 0.05–0.5 Mb (lanes IV). Size standards are *Sc. pombe*, *Pichia*, and *S. cerevisiae* chromosomal DNAs, and bacteriophage λ DNA concatemers.

Table 1. Sizes (Mb) of *Not* I bands detected with the different *Alu* (pX-H, Blur8) and *Kpn* (Tεγ-C, Tεγ-D, ψ^{LTR}) probes

	Clone				
	pX-H	Blur8	Tεγ-C	Tεγ-D	ψLTR
	7	7	7	7	
	5.6	5.6			
	5	5	5	5	
	3.4	3.4	3.4	3.4	3.4
			2.9	2.9	2.9
	2.7	2.7	2.7	2.7	
	2.3	2.3	2.3		
	2.1	2.1	2.1		2.1
	1.7	1.7	1.7	1.7	
	1.4	1.4	1.4		
	1.3	1.3	1.3	1.3	1.3
	1.1	1.1	1.1		
	0.97	0.97			
	0.82	0.82	0.82		
	0.78	0.78	0.78		
	0.75	0.75	0.75		
	0.68	0.68	0.68	0.68	
	0.65	0.65	0.65		
	0.61	0.61			
	0.54	0.54	0.54		
	0.49	0.49	0.49		
	0.45	0.45	0.45		
	0.43	0.43	0.43		
	0.40	0.40	0.40	0.40	
	0.37	0.37	0.37	0.37	
	0.35	0.35	0.35		
	0.34	0.34			
	0.33	0.33	0.32	0.32	
	0.32	0.32	0.27		
	0.27	0.27	0.23	0.23	
	0.23	0.23			
	0.20				
	0.19	0.19			
	0.17	0.17	0.16	0.16	
	0.16	0.16	0.14		
	0.14	0.14			
	0.087				
	0.053	0.053			
	0.043	0.043			
	0.014	0.014			
Total					
Mb	44.4	44.2	38.2	26.1	9.7
No. of bands	39	37	27	13	4
Average					
length, Mb	1.1	1.2	1.4	2	2.4

Clone pX-H contains five different *Alu* sequences. Clone Blur8 contains a single *Alu* sequence. Clone Tεγ-C contains a 1.2-kb 3′-end *Kpn* sequence. Clone Tεγ-D contains a 0.8-kb *Kpn* sequence adjacent to the 5′ end of Tεγ-C. Clone ψ^{LTR} contains 0.45-kb sequences near a *Kpn* 5′ end.

Poisson statistics estimate the probability of not finding an *Alu* sequence in a 0.1-Mb fragment is $<10^{-4}$, while the probability of not finding a 3′-end *Kpn* sequence is 0.036–0.19. Although at least a few of the smaller *Not* I bands should be missing *Kpn* sequences, the number of such bands actually observed is more than expected from Poisson statistics. For instance, for *Kpn*-containing fragments of 0.2 Mb or less, 6.8–7.8 fragments are expected and 2 fragments are observed.

The *Kpn* probe Tεγ-D detected a lower number of *Not* I fragments than Tεγ-C, probably because the internal L1 sequences are less conserved than the 3′-end sequences. Probe Tεγ-D detected only 13 bands, ranging from 0.16 to 7.0 Mb with a total length of 26.1 Mb and an average size of 2.0

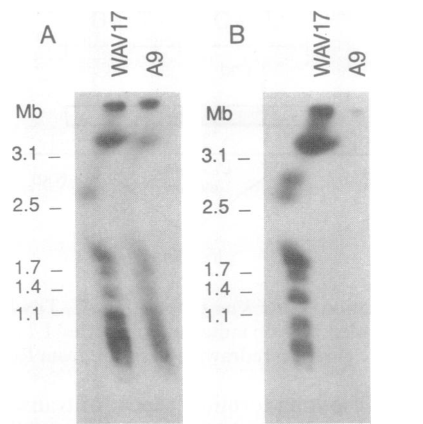

FIG. 3. A 2.9-Mb *Not* I fragment detected with a *Kpn* probe but not with an *Alu* probe. (*A*) Hybridization with the *Alu* probe pX-H. (*B*) The same membrane hybridized with the 3′ *Kpn* probe Tεγ-C. Size standards are *Pichia* and *S. cerevisiae* chromosomal DNAs. The slight hybridization of the *Alu* probe to the A9 sample was probably due to contamination of this particular sample with WAV17 DNA.

Mb (Table 1). The *Kpn* 5′-end probe ψ^{LTR} detected only four *Not* I bands, ranging from 1.3 to 3.4 Mb with a total length of 9.7 Mb and an average size of 2.4 Mb (Table 1). Interestingly, one band (2.1 Mb) was detected with all probes used in this study except probe Tεγ-D. All bands detected by Tεγ-D and ψ^{LTR} were also detected with Tεγ-C. The small number of fragments seen with ψ^{LTR} is in agreement with the expected low frequency of the 5′ end of L1. In all, 40 *Not* I bands were detected with the three *Kpn* and the two *Alu* probes, with a total length of 47.3 Mb and an average size of 1.2 Mb. Twenty-six *Not* I bands (65%) were detected with both families of repeats (*Kpn* and *Alu*). The intensity of hybridization detected in *Not* I bands by using various probes is spread over a wide range. The frequency of *Alu* (Table 2) and *Kpn* (Table 3) sequences, as measured by intensity of hybridization signal per megabase of DNA, varies by >20-fold.

The results shown here reveal a nonrandom distribution of the interspersed repeats in chromosome 21 with a preference for repeats to cluster in specific fragments of the chromosome. For instance, one 2.9-Mb *Not* I fragment did not contain an *Alu* repeat, and 13 *Not* I fragments (from 0.014 to 5.6 Mb, with a total length of 9.1 Mb) did not hybridize to the *Kpn* repeat. Previous studies on *Kpn* or *Alu* repeats also report nonrandom or even complementary distributions. Soriano *et al.* (16), using density-gradient centrifugation, separated four major components of human DNA with a G+C content varying from 37% to 49%. The *Alu* repeat occurred more frequently in the heavy, G+C-rich, components, whereas the *Kpn* repeat was more represented in the light components. This distribution is consistent with the fact that *Alu* and *Kpn* elements contain 56% and 42% G+C, respectively. *In situ* hybridization experiments show prominent clustering of *Alu* and *Kpn* repeats in Giemsa-stained light and dark bands, respectively (17, 18, 37). These results are consistent with the observation that Giemsa light bands are richer in G+C content than Giemsa dark bands (38). Moyzis *et al.* (24) analyzed the distribution of *Alu* repeats appearing in the GenBank sequence data base and observed local domains that they termed rich and poor in *Alu* sequence density. These authors also showed, by *in situ* techniques, that centromeric heterochromatin regions are at least 50-fold underrepresented in *Alu* sequences. The 7.0-Mb chromosome 21 *Not* I fragment is located in the centromeric region (R. Oliva, J.S., H. Ichikawa, M. Murata, M. Ohki, and C.L.S., unpublished data). It has one of the lowest frequencies of *Alu* repeats (see Table 2), with a 2.8-fold (213/75) underrepresentation of *Alu* as compared with the average *Not*

Table 2. Quantitation of *Alu* repeat hybridization signal detected with probe pX-H in *Not* I bands in chromosome 21 from WAV17

Mb	Intensity	Factor	RI	Frequency
7	524	1.00	524	75
5.6	274	1.00	274	49
5	294	1.00	294	59
3.4	751	1.00	750	221
2.7	103	3.20	338	125
2.3	42	3.20	137	59
2.1	125	3.20	410	195
1.7	104	3.20	340	200
1.4	43	3.20	140	100
1.3	56	3.20	184	142
1.1	146	3.20	467	425
0.97	2.5	37.20	94	97
0.82	5.5	37.20	203	248
0.78	15	37.20	557	714
0.75	5.6	37.20	208	277
0.68	3.2	37.20	118	174
0.65	10	37.20	375	577
0.61	4.1	10.60	154	252
0.54	9	10.60	96	178
0.49	7.4	10.60	79	161
0.45	7.5	10.60	80	179
0.43	8	10.60	85	197
0.40	2.6	10.60	28	70
0.37	6.6	10.60	70	189
0.35	12	1.51	18	52
0.34	15	1.51	22	65
0.33	61	1.25	76	228
0.32	44	1.25	56	174
0.27	26	1.25	33	123
0.23	16	1.25	20	87
0.20	25	1.25	31	154
0.19	19	1.25	23	124
0.17	20	1.25	25	142
0.16	40	1.25	50	315
0.14	14	1.25	18	129
0.087	12	1.25	15	173
0.053	8.4	1.25	11	199
0.043	8.4	1.25	11	245
0.014	13	1.25	16	1129
			Average	213

Intensity is the total number of arbitrary counts detected in a band using a PhosphorImager. Factor is the correction used to have the identical number of counts in all the membranes. Fragments 7–3.4 Mb were measured on the same membrane and other results were scaled to this membrane. RI (relative intensity) is the intensity of each band scaled by the factor needed to correct for different total amounts of counts seen in the membranes used. Frequency is the relative intensity per megabase of each band.

I band (Table 2). This is much less than seen by *in situ* methods, but it is still quite significant. In contrast, the 7.0-Mb band has a nearly average frequency of occurrence of *Kpn* repeats (Table 3). Hence, it appears that *Kpn* repeats occur relatively more frequently than *Alu* repeats in the centromere of chromosome 21.

The frequency of appearance of both *Alu* and *Kpn* repeats varies by almost 25-fold in individual *Not* I bands. In principle, such large differences in particular *Not* I bands could be due to comigration of multiple *Not* I fragments in PFGE. Indeed, the 3.4-, 2.1-, and 0.043-Mb *Not* I bands are known to contain more than one *Not* I fragment (refs. 39 and 40; D. Wang, H. Fang, and C.L.S., unpublished data). The 0.043-Mb band contains the ribosomal DNA repeat; it is estimated to consist of at least 13 copies by partial digestion experiments (unpublished data). Each of these other bands contains at least two DNA fragments detected with single copy probes located in different regions of the chromosome. These multiple fragments add an additional 6.1 Mb to the minimum size of chromosome 21. However, it is unlikely that comigration is the major reason for differences in the frequencies of the repeats in the bands, because then the size of chromosome 21 would be much larger than expected from estimates of the total size of the human genome (see below).

Table 3. Quantitation of 3′-end *Kpn* repeat hybridization signal detected with probe Tεγ-C in *Not* I from WAV17 human chromosome 21

Mb	Intensity	Factor	RI	Frequency
7	91	1.00	91	13
5	58	1.00	58	12
3.4	81	1.00	81	24
2.9	10	4.56	47	16
2.7	8.3	4.56	38	14
2.3	3.5	4.56	16	7
2.1	25	4.56	113	54
1.7	9.4	4.56	43	25
1.4	3.8	4.56	17	12
1.3	11	4.56	50	38
1.1	11	4.56	48	44
0.82	0.6	11.40	6.9	8
0.78	2.2	11.40	25	32
0.75	0.5	11.40	6.0	8
0.68	1.3	11.40	15	23
0.65	1.1	11.40	13	20
0.54	0.4	2.66	1.2	2
0.45	0.7	2.66	1.9	4
0.43	1.0	2.66	2.5	6
0.40	1.2	2.66	3.1	8
0.37	2.8	2.66	7.5	20
0.35	0.6	2.66	1.6	4
0.32	3.7	2.66	10	31
0.27	0.8	2.66	2.0	8
0.23	0.6	2.66	1.6	7
0.16	0.6	2.66	1.6	10
0.14	0.7	2.66	1.9	14
			Average	17

See Table 2 for explanation of terms.

The 26 bands that hybridize with both *Alu* and *Kpn* probes represent 65% of the total number of bands. The length of these bands is 35.3 Mb, representing 75% of the total length of the bands. Hence, the data presented here do not support the notion of a purely complementary *Alu* and *Kpn* distribution on human chromosome 21. Extensive overlap of *Alu* and *Kpn* repeats is seen on fragments containing up to 75% of the chromosome length. This overlapping distribution of many *Alu* and *Kpn* repeats is in agreement with evidence that these repeats are associated in human cloned DNA sequences (41, 42). Miyake *et al.* (42) even propose the formation of *Alu*/*Kpn* mobile elements during evolution. *Alu* and *Kpn* sequences appear to be codistributed on a large fraction of chromosome 21, but our results do not indicate whether these sequences are actually near neighbors.

The total apparent length of the *Not* I bands detected with the repeated DNA probes reveals a minimum size of human chromosome 21 of 47.3 or 53.5 Mb when a minimum number of overlapping fragments are considered (see above). This size is larger than the 43 Mb previously determined by summation of 33 independent *Not* I fragments on the q arm detected by using single copy probes (43) and is much larger than that spanned by existing physical maps [e.g., 9 fragments covering 8 Mb (44) or 15 fragments covering 8.9 Mb of the distal long arm (D. Wang, H. Fang, and C.L.S., unpublished results) and 7 fragments spanning 20 Mb of the proximal long arm (39)]. A more accurate size of chromosome 21 awaits further experiments to detect other comigrat-

ing *Not* I bands. There is also a low probability that some additional *Not* I fragments will be found that do not contain *Alu* and *Kpn* interspersed repeats. For example, Poisson statistics estimate that the probability of not finding either an *Alu* or a *Kpn* sequence in a fragment of 0.1 Mb is $<3.7 \times 10^{-6}$. Therefore, such fragments are likely to be very small. The same experimental approach can be applied to estimate a minimum size for other chromosomes cloned in hybrid cell lines. Furthermore, a comparison of fragments detected with single copy sequences with fragments detected by repeats will reveal those fragments that are not yet assigned to single copy sequences (39). Such fragments must be contained in restriction map gaps and need to be the focus of mapping end game strategies.

Could the apparent size of 53.5 Mb for chromosome 21 be an artifact arising from partial digestion? This is unlikely, because most single copy probes and *Not* I linking clones detect single and double *Not* I bands in WAV17, respectively. These results indicate that few if any *Not* I sites are partially methylated. Bird (45) compared the occurrence of restriction enzyme sites with CpGs in CpG islands at 19 human genes. He concluded that *Not* I cuts 3 times more than predicted, in accordance with earlier indications (33). *Not* I usually cuts in CpG islands. Inter-island sites are rare and are usually blocked by methylation. Thus, *Not* I cleavage will seldom be partial, and it is likely that chromosome 21 is really at least 53.5 Mb, 5% larger than previous estimates. It is too early to determine whether the entire human genome is larger than expected or whether chromosome 21 is anomalous.

We thank Drs. Aleksander Milosavljevic, Jerzy Jurka, and Pat Monardo for the use of their human *Alu* sequence identification computer program and Oscar Loureiro for helping us with the statistical analysis. This work was supported by grants from the National Institutes of Health (CA39782), U.S. Department of Energy Contracts DE-AC03-76SF0009 and DEF602-87ER-GD852, and the Ministry of Education and Science of Spain.

1. Korenberg, J. R., Pulst, S. M., Kawashima, H., Ikeuchi, T., Yamamoto, K., Ogasawara, N., Schonberg, S. A., West, R., Kojis, T. & Epstein, C. J. (1989) *Am. J. Hum. Genet.* **43**, A110 (abstr.).
2. McCormick, M. K., Schinzel, A., Petersen, M. B., Stetten, G., Driscoll, D. J., Cantu, E. S., Tranebjaerg, L., Mikkelsen, M., Watkins, P. C. & Antonarakis, S. E. (1989) *Genomics* **5**, 325–331.
3. Rahmani, Z., Blouin, J.-L., Creau-Goldberg, N., Watkins, P. C., Mattel, J.-F., Poissonnier, M., Prieur, M., Chettouh, Z., Nicole, A., Aurias, A., Sinet, P.-M. & Delabar, J.-M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5958–5962.
4. St. George-Hyslop, P. H., Haines, J. L., Farrer, L. A., Polinsky, R., Van Broekhoven, C., Goate, A., Crapper Mclachlan, D. R., Orr, H., Bruni, A. C., Sorbi, S., Rainero, I., Foncin, J.-F., Pollen, D., Cantu, J.-M., Tupler, R., Voskresenskaya, N., Mayeux, R., Growdon, J., Fried, V. A., Myers, R. H., Nee, L., Backhoven, H., Martin, J.-J., Rossor, M., Owen, M. J., Mullan, M., Percy, M. E., Karlinsky, H., Rich, S., Heston, L., Montesi, M., Mortilla, M., Nacmias, N., Gusella, J. F., Hardy, J. A., *et al.* (1990) *Nature (London)* **347**, 194–197.
5. Goate, A. M., Haynes, A. R., Owen, M. J., Farrall, M., James, L. A., Lai, L. Y. C., Mullan, M. J., Roques, P., Rossor, M. N., Williamson, R. & Hardy, J. A. (1989) *Lancet* **i**, 352–355.
6. Potter, H. (1991) *Am. J. Hum. Genet.* **48**, 1192–1200.
7. Siddique, T., Figlewicz, D. A., Pericak-Vance, M. A., Haines, J. L., Rouleau, G., Jeffers, A. J., Sapp, P., Hung, W.-Y., Bebout, J., Mckenna-Yasec, D., Deng, G., Horvitz, H. R., Gusella, J. F., Brown, D. R., Roses, A. D. & Collaborators (1991) *N. Engl. J. Med.* **324**, 1381–1384.
8. Smith, C. L., Lawrance, S. K., Gillespie, G. A., Cantor, C. R., Weissman, S. M. & Collins, F. (1987) *Methods Enzymol.* **151**, 461–489.
9. Cantor, C. R., Pevny, L. & Smith, C. L. (1988) in *Evolutionary Tinkering in Gene Expression*, eds. Grunber-Manago, M., Clarck, B. F. C. & Zachau, H. G. (Plenum, New York), NATO ASI Series, Vol. 169, pp. 143–154.
10. Korenberg, J. R. & Engels, W. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3382–3386.
11. Morton, N. E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7474–7476.
12. Schwartz, D. C., Saffran, W., Welsh, J., Haas, R., Goldenberg, M. & Cantor, C. R. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 189–195.
13. Schwartz, D. C. & Cantor, C. R. (1984) *Cell* **37**, 67–75.
14. Singer, M. F. (1982) *Int. Rev. Cytol.* **76**, 67–112.
15. Skowronski, J. & Singer, M. F. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51**, 457–464.
16. Soriano, P., Meunier-Rotival, M. & Bernardi, G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1816–1820.
17. Korenberg, J. R. & Rykowski, M. C. (1988) *Cell* **53**, 391–400.
18. Chen, T. L. & Manuelidis, L. (1989) *Chromosoma* **98**, 309–316.
19. Jurka, J. (1990) *Nucleic Acids Res.* **18**, 137–141.
20. Tashima, M., Calabreta, B., Torelli, G., Scofield, M., Maizel, A. & Saunders, G. F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1508–1512.
21. Rinehart, F. P. (1981) *Biochemistry* **20**, 3003–3010.
22. Hwu, H. R., Roberts, J. W., Davidson, E. H. & Britten, R. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3875–3879.
23. Jelinek, W. R. & Schmid, C. W. (1982) *Annu. Rev. Biochem.* **51**, 813–844.
24. Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J.-R., Burks, C., Sirotkin, K. M. & Goad, W. B. (1989) *Genomics* **4**, 273–289.
25. Grimaldi, G., Skowronski, J. & Singer, M. F. (1984) *EMBO J.* **3**, 1753–1759.
26. Raziuddin, A., Sarkar, F. H., Dutkowski, R., Shulman, L., Ruddle, F. H. & Gupta, S. L. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5504–5508.
27. Smith, C. L., Kico, S. R. & Cantor, C. R. (1988) in *Genome Analysis: A Practical Approach*, ed. Davis, K. (IRL, Oxford), pp. 41–72.
28. Fang, H., Zhang, T.-Y., Oliva, R., Fan, J.-B., Nicolik, J. & Smith, C. L. (1991) in *Mapping the Human Genome*, ed. Naylor, S. (Academic, San Diego), in press.
29. Mathew, M. K., Smith, C. L. & Cantor, C. R. (1988) *Biochemistry* **27**, 2904–1910.
30. Fan, J.-F., Chikashige, Y., Smith, C. L., Niwa, O., Yanagida, M. & Cantor, C. R. (1989) *Nucleic Acids Res.* **17**, 2801–2818.
31. Perlino, E., Paonessa, G. & Ciliberto, G. (1985) *Nucleic Acids Res.* **13**, 8359–8377.
32. Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. & Schmid, C. W. (1981) *J. Mol. Biol.* **151**, 17–33.
33. Smith, C. L., Lawrance, S. K., Gillespie, G. A., Cantor, C. R., Weissman, S. M. & Collins, F. S. (1987) *Methods Enzymol.* **151**, 461–489.
34. Jurka, J. & Milosavjlevic, A. (1991) *J. Mol. Evol.* **75**, 105–121.
35. Rogan, P. K., Pan, J. & Weissman, S. M. (1987) *Mol. Biol. Evol.* **4**, 327–342.
36. Feinberg, A. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
37. Filatov, L. V., Mamayeva, S. E. & Tomilin, N. V. (1987) *Mol. Biol. Reports* **12**, 117–121.
38. Holmquist, G. P. (1988) in *Chromosomes and Chromatin*, ed. Adolph, K. W. (CRC, Boca Raton, FL), pp. 76–121.
39. Ichikawa, H., Shimizu, K., Saito, A., Wang, D., Smith, C. L., Cantor, C. R., Kobayashi, H., Kaneko, Y., Miyoshi, H. & Okhi, M. (1991) *Proc. Natl. Acad. Sci. USA* **89**, 23–27.
40. Saito, A., Abad, J. P., Wang, D., Ohki, M., Cantor, C. R. & Smith, C. L. (1991) *Genomics* **10**, 618–630.
41. Shafit-Zagardo, B., Maio, J. & Brown, F. L. (1982) *Nucleic Acids Res.* **10**, 3175–3193.
42. Miyake, T., Migita, K. & Sakaki, Y. (1983) *Nucleic Acids Res.* **11**, 6837–6846.
43. Gardiner, K., Horisberger, M., Kraus, J., Tantravahi, U., Korenberg, J., Rao, V., Reddy, S. & Patterson, D. (1990) *EMBO J.* **9**, 25–34.
44. Burmeister, M., Kim, S., Price, E. R., De Lange, T., Tantravani, U., Myers, R. M. & Cox, D. R. (1991) *Genomics* **9**, 19–30.
45. Bird, A. P. (1989) *Nucleic Acids Res.* **17**, 9485.