© 1992 Oxford University Press
Nucleic Acids Research, 1992, Vol. 20, No. 22 6033–6042

PCR amplification of tandemly repeated DNA: analysis of intra- and interchromosomal sequence variation and homologous unequal crossing-over in human alpha satellite DNA

Peter E.Warburton[1,2] and Huntington F.Willard[1,*]
[1]Department of Genetics, Stanford University, Stanford, CA 94305, USA and [2]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Received July 9, 1992; Revised and Accepted October 13, 1992

ABSTRACT

Tandemly repeated DNA can comprise several percent of total genomic DNA in complex organisms and, in some instances, may play a role in chromosome structure or function. Alpha satellite DNA is the major family of tandemly repeated DNA found at the centromeres of all human and primate chromosomes. Each centromere is characterized by a large contiguous array of up to several thousand kb which can contain several thousand highly homogeneous repeat units. By using a novel application of the polymerase chain reaction (repPCR), we are able to amplify a representative sampling of multiple repetitive units simultaneously, allowing rapid analysis of chromosomal subsets. Direct sequence analysis of repPCR amplified alpha satellite from chromosomes 17 and X reveals positions of sequence heterogeneity as two bands at a single nucleotide position on a sequencing ladder. The use of TdT in the sequencing reactions greatly reduces the background associated with polymerase pauses and stops, allowing visualization of heterogeneous bases found in as little as 10% of the repeat units. Confirmation of these heterogeneous positions was obtained by comparison to the sequence of multiple individual cloned copies obtained both by PCR and non-PCR based methods. PCR amplification of alpha satellite can also reveal multiple repeat units which differ in size. Analysis of repPCR products from chromosome 17 and X allows rapid determination of the molecular basis of these repeat unit length variants, which appear to be a result of unequal crossing-over. The application of repPCR to the study of tandemly repeated DNA should allow indepth analysis of intra- and interchromosomal variation and unequal crossing-over, thus providing insight into the biology and genetics of these large families of DNA.

INTRODUCTION

The genomes of most complex eukaryotes contain large amounts of tandemly repeated DNA, which comprise several percent of the total genetic material in some organisms. These DNA families consist of particular repeat units that are organized in a head-to-tail fashion, forming arrays as large as several thousand kb. Although no function has yet to be definitively attributed to such non-coding tandemly repeated DNA families, they are often located in functionally provocative places on chromosomes, such as the centromeres, telomeres and nucleolar organizing regions (1). A curious property of tandemly repeated DNA families is the high degree of sequence homogeneity seen between repeat units within a species. This is thought to be the consequence of a concerted process of homogenization within and between chromosomes, presumably by the complex interaction of genomic turnover processes such as unequal crossing-over and sequence conversion (2, 3).

One such family of tandemly repeated DNA is alpha satellite DNA, which is the major class of DNA located at the centromeres of all human and primate chromosomes (4). The most basic repeat unit of alpha satellite DNA consists of a ~171 bp monomer. Distinct linear arrangements of these monomers form chromosome-specific higher-order repeat units, which are in turn tandemly repeated to form arrays as large as several thousand kb. Although individual monomers can be 35% diverged from each other, higher-order repeat units from a particular chromosomal subset are in general greater than 95% identical; it is this high degree of similarity between multiple repeat units, as well as their high copy number, that makes them useful as chromosome-specific hybridization probes (5, 6). For example, the centromeric region of the X chromosome is characterized by approximately 1500 copies of a 2.0 kb repeat unit, which is composed of 12×171 bp monomers (12mer) (7, 8). Similarly, the centromeric region of chromosome 17 consists of approximately 1000 copies of a 2.7 kb higher order repeat unit, which is composed of 16×171 bp monomers (16mer), and related variants (9, 10). Other less abundant but closely related repeat

* To whom correspondence should be addressed at: Department of Genetics, School of Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA

units that consist of different integral numbers of monomers are also present on these chromosomes; such variants are presumably generated by unequal crossing over (10−12). Such repeat units can be used as polymorphic markers in the genetic analysis of centromeric loci (13).

The polymerase chain reaction (PCR) has proven useful in the study of alpha satellite DNA. By taking advantage of DNA sequence information within and between chromosomal subsets, PCR primer sets have been designed that amplify alpha satellite DNA in a chromosome-specific manner from at least six different subsets of this DNA family (14−16). The resulting PCR products can be used as chromosome-specific *in situ* hybridization probes and in some cases reveal polymorphic repeat units that can be used for rapid genotyping (14). In this report, we further apply the PCR reaction to the analysis of the organization and evolution of tandemly repeated DNA.

In principle, annealing sites for unbiased PCR primers targeted to tandemly repeated DNA should be found in most or all repeat units in the array and, therefore, should amplify multiple repeat units simultaneously. In a technique we have called repPCR, we show that PCR of tandemly repeated DNA amplifies multiple repetitive units that are representative of the genome. Direct sequencing of repPCR products shows positions of sequence heterogeneity between repeat units as two bands in the same base position on the sequencing ladder. This extends the widely used method of directly sequencing PCR products from single-copy genomic DNA, where heterozygosity for a mutation is revealed as two bands on a sequencing ladder (e.g., 17). In repPCR, however, instead of examining just two alleles, large numbers of repeat units are examined simultaneously by PCR.

When PCR primers amplify multiple repeat units of different sizes, they can be directly sequenced to determine the molecular basis of the polymorphism. In addition, multiple independent clonal copies of these repeat units can be easily obtained from the repPCR products, which can be sequenced independently. The data obtained by repPCR are compared and contrasted to previously published reports of sequence variation and unequal crossing-over from the X chromosome (18) and from chromosome 17 (11), respectively, which were based on analysis of a limited number of cloned repeat units obtained by conventional, non-PCR based methods. RepPCR can be used to greatly facilitate the study of repeat unit variation in tandemly repeated DNA and thus should prove valuable for evolutionary and functional studies of this abundant class of DNA.

MATERIALS AND METHODS

Primer selection and PCR conditions

PCR and sequencing primers of approximately 20 bp were chosen from the published sequence of chromosome-specific higher-order repeat units from either chromosome 17 (10) or the X chromosome (7). Primers 17-1A, 17-2A, X-3A, X-4A and their PCR conditions have been previously described (14). PCR reaction conditions for additional primers used in this study are the same as described for other alpha satellite primers (14) and were used for PCR amplification at 30 cycles. Primers used for sequencing reactions were always HPLC purified (Operon Technologies, Alameda, CA), which produced cleaner sequencing ladders, presumably due to complete removal of shorter length oligonucleotide contaminants. Sequencing primers were not always rigorously determined to be chromosome-specific, since they were used only to sequence either purified PCR amplified DNA or plasmid DNA. Additional primers used in this study for both PCR and sequencing, and their positions in the published clones, are as follows:

Chromosome X alpha satellite primers (7):

X-5A	5′-ATT CTC AGA AAG TTT TCT GC-3′	pos. 758−777
X-5B	5′-ATA ACG AAT GTT CAG CTC CC-3′	pos. 477−458
X-8E	5′-CTG AGA ATG CTT CCG TTT GCC-3′	pos. 1278−1258
X-3B	5′-TGT GTT TAG TTA TGG GAA ATT AT-3′	pos. 1435−1413
X-4B	5′-AAG CCT TTT CCT TTA TCT TCA CA-3′	pos. 1923−1946

Primers X-3B and X-4B are reverse complements of primers X-3A and X-4A, respectively.

Chromosome 17 alpha satellite primers (10):

17-7B3	5′-GCT TCT GTC TAG ATT TTA TGC-3′	pos. 1063−1043
17-16E3	5′-CTT TCT CAG AAA ATT CTT TGG-3′	pos. 2589−2609
17-9D	5′-ATG GTC ACA TAA AAA CTG GAG-3′	pos. 1379−1399
17-12E2	5′-AAA CTC TTG GTG ACG ACT GAG-3′	pos. 1927−1947
17-1ASalI	5′-(CAG TCA GTC GAC) TGT TTA GTC AGC TGA AAT T-3′	pos. 38−20
17-2ABamHI	5′-(CCT GAG GAT CC)A TAA CTG CAC CTA ACT AAA CG-3′	pos. 868−888

For the latter two primers, the region in parentheses consists of additional non-alpha satellite sequence containing restriction enzyme recognition sites used for cloning (underlined).

Somatic cell hybrids

Construction and characterization of rodent/human somatic cell hybrids have been described elsewhere (6, 13). Hybrid 1 (L65-14A) and Hybrid 2 (LT23-4C) have previously been used for the analysis of chromosome 17 alpha satellite (9, 19). Hybrids containing one or more human X chromosomes (t48-1ac1a) and hybrids containing a single human X chromosome as their only human complement (t60-12, AHA11a, A2-4) have been described (20).

Lambda clones 25 and 72, described elsewhere (18), contain only tandem copies of X chromosome alpha satellite 2.0 kb repeat units, obtained from a partial Sau 3A1 library of hybrid A2-4. They were characterized by restriction mapping and Southern blotting (S.Durfy, unpublished).

Cloning and sequencing of PCR amplified DNA

PCR amplified DNA was phenol extracted and ethanol precipitated prior to restriction enzyme digestion. Chromosome 17 alpha satellite DNA PCR amplified with primers 17-1A and 17-2A was digested with either Xba I/Eco RI or Xba I/Pvu II (see Fig. 1), ligated with similarly digested pUC 18, and transformed into E.coli strain DH5α. The recognition site for Pvu II is incorporated into the sequence of primer 17-1A (14). Alternatively, primers 17-1ASalI and 17-2ABamHI, which have Sal I and Bam HI recognition sequences attached 5′ to primers 17-1A and 17-2A, respectively (see above), were used to PCR amplify chromosome 17 alpha satellite DNA. Amplified DNA was digested with Bam HI and Sal I, and ligated into similarly digested pUC 19. In each case, multiple clones were obtained, representing the different size higher-order repeat units amplified (14), which were selected by size of their inserts for further analysis. Clones were sequenced with Sequenase (USBiochemicals) with modifications described below, using either primers 17-1A (when present in clone) or M13 reverse and/or M13−20 primers (Operon Technologies). Additional sequence was obtained with primers 17-9D and 17-12E2 (see

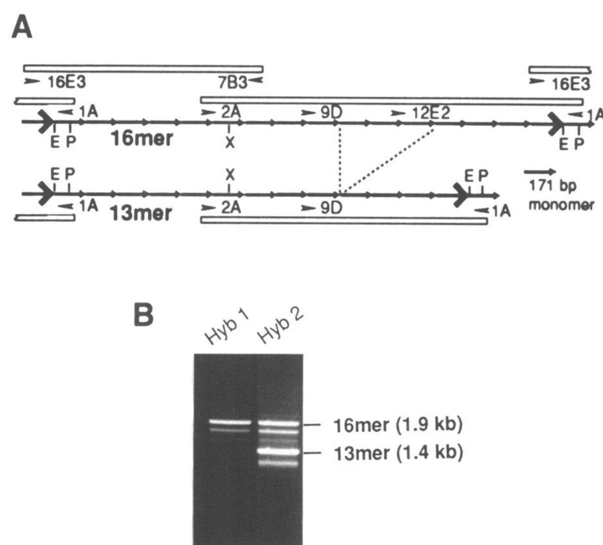

Figure 1. RepPCR of Chromosome 17 Alpha Satellite DNA. **A)** Schematic of the PCR amplification of 16mer and 13mer repeat units. A 1.9 kb band is amplified from the 16mer with primers 17-1A and 17-2A, while a 1.4 kb band is amplified from the 13mer, due to a three monomer (~500 bp) deletion relative to the 16mer (14). Primers 17-7B3 and 17-16E3 amplify a 1.2 kb band from both the 16mer and the 13mer. Due to the large number of tandem repeat units most or all of which have sites for these primers, multiple repeat units are amplified simultaneously. Primers 17-9D and 17-12E2 are used as sequencing primers on gel purified 1.9 kb and 1.4 kb PCR products. **B)** Ethidium bromide stained agarose gel of PCR amplified chromosome 17 alpha satellite DNA. Hybrid 1 and 2 are somatic cell hybrids that each contain a different single chromosome 17. The other bands represent other minor repeat unit length variants (15mer, 14mer, and 12mer) characteristic of this alpha satellite subset (14). E-EcoRI, P-Pvu II.

Fig. 1). PCR amplifying and sequencing of clones and mixtures of clones, used as control experiments, was performed by PCR amplifying approximately 1 ng (total) linearized plasmid DNA with primers 17-1A and 17-2A, gel purifying and sequencing as described below.

Direct sequencing of PCR products

Approximately three to five PCR reactions for each preparation were pooled and ethanol precipitated. The resulting amplified DNA (1–2 μg) was electrophoresed on thin (approximately 0.5 cm) low melting point agarose gels (FMC Seaplaque). The bands of interest were excised in approximately 300 μl agarose, melted at 68°C for 10 min, and cooled to 42°C. 20 units of β-agarase (Calbiochem) were added, and incubated at 37°C for two hours with agitation at 20 min intervals. Mixtures were placed on ice for 15 min, and micro-centrifuged for 10 min to remove any undigested agarose. Supernatants were ethanol precipitated and resuspended in 10:1 TE (10mM Tris, pH 7.4, 1mM EDTA).

Double strand sequencing was performed using the Sequenase kit (USBiochemicals) with the following specific protocol modifications. Denaturing and annealing was accomplished by placing approximately 1 μg of gel purified PCR amplified DNA with 0.01μg of primer in a boiling water bath for 2 min and placing immediately on ice. Extensions were performed at room temperature for 90 sec in 1× Sequenase reaction buffer as prescribed in the Sequenase protocol, using undiluted (5×) Sequenase labeling mix plus addition of Sequenase $MnCl_2$ buffer. Terminations were performed at 45°C for 90 sec as prescribed, using a 2:1 ratio of each Sequenase ddNTP termination mix to Sequenase extension buffer. Finally, 1 μl of a solution containing 1mM GTP, ATP, TTP, CTP and 2 u/μl Terminal deoxynucleotidyl Transferase (TdT) (USBiochemicals) was added to each termination reaction and incubated at 37°C for 0.5–1 hr (21). Reactions were stopped by addition of 4μl Sequenase stop buffer and placed on ice.

Table 1. Sequence differences within and between alpha satellite repeat units from two chromosomes 17 detected by repPCR

Heterogeneous Positions by repPCR[a]	2525	2532	2542	2543	2583	2630	2647	
Hybrid 1								
repPCR	A/C	T/Δ	C/A	C/G	A/(G)	T/(G)	G/(A)	
proportion in clones[b]								haplotype
4/12[c]	C	Δ	A	G	A	T	G	a
3/12[c]	A	Δ	A	G	A	T	G	b
3/12	A	T	C	C	G	G	A	c
2/12	C	T	C	C	A	T	A	d
total	A(6)/C(6)	T(5)/Δ(7)	C(5)/A(7)	C(5)/G(7)	A(9)/G(3)	T(9)/G(3)	G(5)/A(7)	
Hybrid 2								
repPCR	A	T	C	C	(A)/G	T/G	A	
proportion in clones[b]								haplotype
6/10	A	T	C	C	G	G	A	c
3/10	A	T	C	C	A	T	A	e
1/10	A	T	C	C	G	T	A	f
total	A(10)	T(10)	C(10)	C(10)	A(3)/G(7)	T(4)/G(6)	A(10)	

[a] 7 heterogeneous positions were seen in hybrid 1, while only 2 were seen in hybrid 2. Δ = one basepair deletion. Sequence positions as numbered in Waye & Willard, 1987 (10).
[b] multiple independent 16 mers were cloned from the repPCR reaction, and sequenced with primer 17-1A (see text).
Hybrid 1- 12 independent clones total, Hybrid 2- 10 independent clones total.
[c] From hybrid 1, 5 additional clones were sequenced, but were not independent by the criteria described in text. 4 were of haplotype b, 1 of haplotype a.
Parentheses enclose the less abundant base (<20% estimated).

RESULTS

RepPCR: amplification and examination of multiple repeat units simultaneously

The validity of the representative nature of repPCR was tested by examination of the 2.7 kb 16mer higher-order repeat unit from chromosome 17. Primers 17-1A and 17-2A were used to amplify specifically chromosome 17 alpha satellite from human genomic DNA, giving a 1.9 kb band that corresponds to the 16mer (Fig. 1A,B) (14). DNAs from two mouse/human somatic cell hybrids that each contain a different single chromosome 17 were PCR amplified, the 1.9 kb products gel-purified and ~225 bp directly sequenced (positions 2463 to 2688, as numbered in (10)) using primer 17-1A again as the sequencing primer. Within this region, seven sequence differences were seen within and between the two copies of chromosome 17 (Table 1). Portions of these

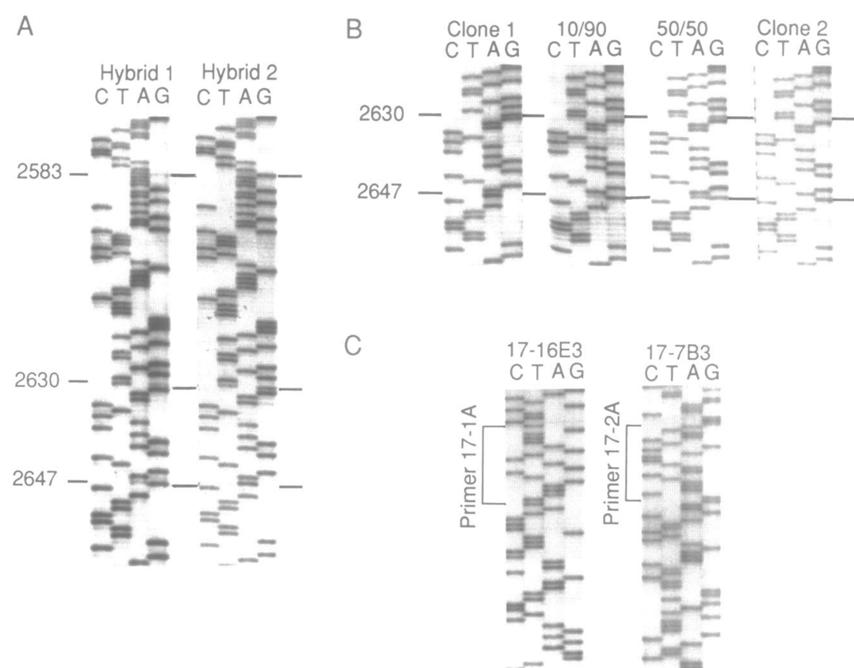

Figure 2. Direct Sequencing of PCR Amplified Chromosome 17 Alpha Satellite DNA. **A**) The 1.9 kb PCR products (16mer) from hybrids 1 and 2 (see Fig. 1B) were directly sequenced using primer 17-1A. Position 2583 shows a heterogeneous A/G; position 2630 shows T/G. Note the different relative intensities of these nucleotides in the two hybrids (see Table 1). Position 2647 shows a heterogeneous A/G only in hybrid 1, with an unambiguous A in hybrid 2 (see Table 1). **B**) DNA from two repeat unit clones were PCR amplified with primers 17-1A ad 17-2A individually and in a series of mixtures of different molar ratios. The resulting 1.9 kb fragments were directly sequenced using primer 17-1A. The individual clones show no heterogeneous positions (Clone 1 and Clone 2). The molar ratios of two mixtures (10:90 and 50:50) are reflected in the relative intensity of the heterogeneous bands, although the 10:90 was occasionally not obvious over background. C) Sequencing of 1.2 kb PCR product amplified and sequenced using primers 17-16E3 and 17B3. Sequencing with primer 17-16E3 allows direct sequencing of the site for primer 17-1A. Sequencing with primer 17-7B3 allows direct sequencing of the site for primer 17-2A. No heterogeneous positions are seen in the sequence of these primers for either hybrid 1 or 2.

sequencing ladders that include three such positions are shown in Fig. 2A. Within hybrid 1, all seven positions were heterogeneous. Six correspond to single basepair changes revealed as two bands in the same position on the sequencing ladder; the seventh position (2532) represents a one basepair deletion. Hybrid 2 showed heterogeneous patterns at only two of the seven positions identified in hybrid 1 (2583 and 2630), the other five showing only an unambiguous single nucleotide. No other heterogeneous positions were detected in hybrid 2 in the 225 bp examined.

A number of findings indicate that the relative intensity of bands at such heterogeneous positions approximates the relative proportion of repeat units containing one or the other base. For example, one of these heterogeneous positions (2525) is found within a previously described polymorphic Sau3A1 site (9). Based on relative intensities of restriction fragments on Southern blots, this site was found in approximately 30−50% of the repeat units in hybrid 1 (see (9)), but was not detectable in hybrid 2 (data not shown). This is consistent with the relative intensites of the bases at this position on the repPCR sequencing ladder, which was estimated to be slightly below 50% C (giving the Sau3A1 site) in hybrid 1, but was judged to be unambiguously (100%) A (abolishing the site) in hybrid 2 (data not shown). Relative intensity differences between bands or between the hybrids were generally reproducible in independent experiments. Thus, the heterogeneous bases at positions 2542 and 2543 appeared to be of approximately equal intensity in hybrid 1 (data not shown), suggesting approximately equal distribution of these positions within the repeat units of that hybrid. On the other hand, positions 2583 and 2630, heterogeneous in both hybrids, have different relative intensities in each hybrid (Fig. 2A). As well, at position 2647, the band that is fainter in hybrid 1 is the only band seen in hybrid 2. Where detected, such relative intensity differences are indicated in Table 1 by placing the less abundant base in parentheses. Bands of approximately equal intensity have no parentheses.

The sequencing ladders in Fig. 1A also show some faint bands that are not scored as heterogeneous positions. Most of these faint bands, which are seen in all four lanes of the sequencing ladder, are attributed to polymerase pauses and stops. Such bands are also seen in sequencing of PCR amplified clonal DNA (Fig. 2B). Such pauses, which would interfere with the detection of heterogeneous positions, were greatly reduced in intensity by the addition of TdT to the sequencing reaction (21) (see materials and methods). Other faint bands seen in single additional lanes at some base positions were not reproducible and were, therefore, not included in this analysis. It is possible that these positions may, in some cases, represent low levels of sequence heterogeneity between repeat units. As such, heterogeneous positions were scored conservatively, where only those positions that were reproducible multiple times from independent PCR and sequencing reactions were considered. This may result in an underestimation of overall sequence heterogeneity.

In order to further assess the validity of repPCR, a total of 22 independent plasmids containing most or all of the 1.9 kb alpha satellite fragment as insert were cloned from the PCR products,

using several different cloning strategies (see materials and methods). Twelve clones were obtained from hybrid 1, ten from hybrid 2. These clones were judged to be independent either because they are from different hybrids, or because of unique differences in the sequence relative to all other clones (within the 225 bp region discussed here or in other regions within the 1.9 kb insert). An additional five clones were obtained from hybrid 1 that were not found to be independent over the 225 bp sequenced; however, further sequence analysis of other regions of the 1.9 kb inserts might establish their independence as well. The fact that at least 22 out of 27 clones obtained from repPCR were independent supports the tenet that the PCR reaction is amplifying multiple repeat units simultaneously.

Table 1 shows the distribution of bases at these heterogeneous positions in the sets of clones from each hybrid, separating the different clones into similar patterns or haplotypes, and totaling the proportion of nucleotides at each position for each hybrid. In most instances (position 2647 being an apparent exception), the relative abundance of heterogeneous positions in this region in each set of clones is consistent with the positions and relative intensities seen in the repPCR-derived direct sequence data. However, many more clones would be required to represent a statistically significant estimate of the variation within the approximately 1000 repeat units found in each hybrid. The general conclusion is supported by the relative intensities of positions 2583 and 2630 between the two hybrids (Fig. 2A and Table 1) and by the unambiguous base positions scored in hybrid 2, each of which showed the same base in 10 out of 10 clones. In hybrid 1, five other changes, undetected by repPCR, were seen in 2 out of the 12 clones examined (data not shown). This may represent a threshold (10–20%) on the ability to visualize low frequency sequence variation at heterogeneous positions with this technique, given the conservative scoring.

PCR amplification of cloned alpha satellite repeat units

In order to further assess the validity of these data, a number of controls were performed. Two clones were selected that contained internal priming sites for primers 17-1A and 17-2A (see materials and methods) and between them had all seven of the differences seen between the hybrids in the repPCR analysis (haplotype a and c, Table 1). DNA from these clones, as well as a series of mixtures of different molar ratios (from 10:90 to 50:50), were subjected to PCR amplification with primers 17-1A and 17-2A. The resulting 1.9 kb PCR products were gel purified and the same 225 bp directly sequenced using primer 17-1A as sequencing primer. Fig. 2B shows portions of the resulting sequencing ladders, including positions 2630 and 2647. As expected, the sequence of the PCR amplified pure clones did not show heterogeneity at any position. Moreover, in the PCR amplified mixtures, the relative intensities of the bands reflected the initial molar ratios of the two clones, although the 10:90 ratio was occasionally not detectable over background (Fig. 2B). This again may represent a threshold of detection. Thus, the same primers and identical heterogeneous positions were examined artificially, and were completely consistent with the interpretation from the experimental data.

Since infidelity of DNA polymerase could contribute to the estimate of sequence heterogeneity, the misincorporation rate of the PCR and sequencing reactions were assessed for the particular conditions used here. A single cloned chromosome 17 alpha satellite repeat unit (10) was PCR amplified and seven individual clones obtained. Sequence analysis of these clones in the same region as examined above revealed four base substitutions out of a total of 2037 bp sequenced, giving an estimated misincorporation rate for both Taq polymerase and Sequenase (USB) of ~0.2%, consistent with previous estimates (22). These base substitutions were not at any of the heterogeneous positions described (Table 1), and the detected error rate is well below the level of variation seen between repeat units. Additionally, direct sequence analysis of the PCR amplified clonal mixtures (total of six) showed no additional heterogeneous positions than were present in the original mixtures (six mixtures, 225 bp each) (Fig. 2A). It is formally possible, however, that on rare occasions a cloned repeat unit could be deemed independent due to a misincorporation artifact.

Confirmation of the representative nature of PCR primers

An important caveat in the application of PCR amplification to the study of tandemly repeated DNA is the possibility of primer bias. If a primer was chosen such that it annealed only to a particular subset of repeat units (due to, for example, a heterogeneous position near the 3′ end of the primer), then only the group of repeats containing the heterogeneous position would be amplified. Only base changes specific to this group (e.g. those in linkage disequilibrium (18)) would be detected in the direct sequence analysis. Use of such a biased primer pair could, therefore, result in amplification of a non-representative subset of repeat units. Such a potential bias would be similar to that introduced in many previous studies of satellite DNA, in which clones were often isolated by making use of restriction sites limited to only a subset of the repeat units, as seen on Southern blots (e.g., 18).

To rule out significant primer bias for primers 17-1A and 17-2A from chromosome 17 alpha satellite, an independent set of PCR primers was designed, 17-7B3 and 17-16E3, that amplify a 1.2 kb fragment that contains the sites for primers 17-1A and 17-2A (Fig. 1A). It should be pointed out that this 1.2 kb fragment is amplified from all of the repeat unit length variants seen, because deletions giving rise to these variants do not fall between primers 17-7B3 and 17-16E3 (see (14)). This fragment was directly sequenced from both ends using the PCR primers. Thus, the region of the repeat unit containing the site for primers 17-1A and 17-2A was directly sequenced by this procedure (Fig. 1A) in order to determine if any positions of heterogeneity exist in these primers that would bias their use in repPCR. As can be seen in Fig. 2C, the sequence of these primers contains no heterogeneous base positions and are thus not likely to PCR amplify a biased subset of repeat units for chromosome 17 alpha satellite. This was true for both hybrid 1 and hybrid 2. In the complementary experiment, there were no ambiguous base positions seen in either hybrid in the sequence of primers 17-7B3 and 17-16E3 when directly sequenced from PCR amplified alpha satellite using primers 17-1A and 17-2A (data not shown). This experiment also allowed sequencing of both complementary strands of DNA in the region between the primers, which showed complete agreement and no additional heterogeneous positions in this limited region.

Comparison of repPCR and multiple clone derived sequence

In order to compare data on sequence variation obtained from repPCR to that obtained previously by sequencing multiple single clones obtained by non-PCR based methods (18), PCR amplified alpha satellite DNA from the X chromosome was directly sequenced. DNAs from three mouse/human somatic cell hybrids

Table 2. Comparison of sequence differences in alpha satellite repeat units from the X chromosome seen by repPCR and non-PCR based methods

	heterogeneous positions in clones[a]						
	164	194	244	306	427	453	501
clones[b] (n = 34)	G(26)/T(8)	C(17)/G(17)	A(22)/G(12)	G(22)/T(12)	T(18)/C(16)	A(22)/G(12)	C(25)/G(9)
repPCR[c]	G/T	C/G	A	G	T/C	A/G	C/G

[a]All recurrent heterogeneous positions between PCR primers 3A and 4A that were in >7/34 (>20%) of clones, from ref. 18.
[b]34 independent 2kb alpha satellite repeat unit clones were obtained from at least five X chromosomes, and sequenced (18).
[c]Heterogeneity (if any) seen at these positions in repPCR sequence. Positions 244 and 306 showed no heterogeneity in repPCR analysis (see text for discussion).

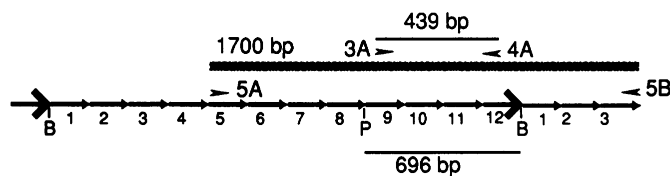

Figure 3. RepPCR Amplification of X Chromosome Alpha Satellite. Primers X-5A and X-5B amplify a 1700 bp fragment from the predominant 2kb 12monomer X chromosome alpha satellite repeat unit. This fragment was subsequently directly sequenced using primers X-3A and X-4A. This repPCR sequence was compared to the overlapping portion of the sequences of the 696 bp Bam HI/Pst I fragment from 34 independent clones (18). P-Pst I, B-Bam HI

that each contain a single X chromosome were amplified using primers X-5A and X-5B, which amplify a 1700bp fragment that corresponds to the predominant repeat on the X chromosome, a 2.0 kb 12mer (Fig. 3). The 1700 bp fragments from the three hybrids were gel purified and sequenced directly, using the primers X-3A and X-4A as sequencing primers (Fig. 3). This 439 bp region of the chromosome X repeat unit found between primers X-3A and X-4A was included in a previous study of the sequence variation found in this alpha satellite subset (18). In that study, sequence analysis was performed on cloned 696 bp Bam HI/Pst I restriction fragments (Fig. 3) from 34 independent 2.0 kb 12mer repeat units, representing at least five different X chromosomes. One of the hybrids (A2-4) used here for PCR analysis was also used previously as a source for cloned repeats (18). Previously, 22 recurrent heterogeneous positions were seen within the set of 34 sequenced clones in the 439 bp region between primers 3A and 4A (Fig. 3). The least frequent bases were detected in from 6% (2 clones) to 50% (17 clones) in the total set of 34 clones (18).

Table 2 lists the position and nucleotides seen at the seven most recurrent heterogeneous positions seen in the 34 clones previously sequenced (18). Examination of the sequencing ladder derived from direct sequencing of the 439 bp repPCR product from each of the three hybrids revealed five base positions where two bands were seen, which correspond to five of these high-copy recurrent heterogeneous positions. No other heterogeneous positions and no differences between the hybrids were seen in the region sequenced. Positions 244 (12/34) and 306 (12/34) were not detectably heterogeneous by repPCR (Table 2). These positions were in complete linkage disequilibrium in the 34 original clones (18), and thus may represent a biased subset of repeat units. This bias could be introduced in two ways, either due to the sequence of the PCR or sequencing primers, or due to the requirement of both a Bam HI and Pst I site in each clone used in the previous study (18). Direct sequencing of repPCR-amplified alpha satellite DNA does reveal five out of seven highly recurring positions of sequence heterogeneity previously detected in a large group of independent clones obtained by conventional non-PCR based methods. However, heterogeneous positions seen previously in less than 20% of the cloned repeat units were not detected, which again may represent a threshold of detection for repPCR.

RepPCR analysis of a product of unequal crossing-over

Alpha satellite DNA from chromosome 17 is predominantly characterized by a 2.7 kb 16mer, seen on all chromosomes 17 examined (10). An additional higher-order repeat unit that consists of 13 monomers is also seen on 35% of chromosomes 17 examined, which forms the basis for haplotypic analysis of this subset (19). The molecular basis for the 13mer has been previously determined by comparison of the sequence of a single cloned copy of the 16mer and 13mer and has been described as an unequal crossing-over between two misaligned 16mers, resulting in a three monomer deletion (11). As seen in Fig. 1A,B, alpha satellite primers 17-1A and 17-2A amplify both types of higher-order repeat units in hybrid 2, resulting in 1.9 kb and 1.4 kb PCR products that correspond to the 16mer and 13mer, respectively. The chromosome 17 in hybrid 1, however, does not show the 1.4 kb band, and is therefore of the other major haplotype (14, 19).

The 1.9 kb band (corresponding to the 16mers) and the 1.4 kb band (corresponding to the 13mers) from hybrid 2 were PCR amplified and gel purified. Approximately 200 bp were sequenced directly from each using both primer 17-9D (16mer and 13mer) and 17-12E2 (16mer only). These primers were positioned to allow analysis across the deletion in the 13mer and corresponding regions in the 16mers (see Fig. 1A). By aligning the sequence of the 13mer to the sequence of the 16mer in both registers, an analysis of the putative misalignment and crossing-over between two 16mers could be performed (Fig. 4) in a similar way to that previously done with single clones (11). The sequence obtained with repPCR, however, should be more representative of the overall repeat units than the single clones used for the previous analysis. As well, the 16mers from the same chromosome 17 as the 13mers are examined in this study.

The sequence differences and heterogeneous positions seen in the repPCR sequence of the 16mers and 13mers (Fig. 4) are of interest since some of them are located at positions that were different in the comparison of the two cloned copies (11). For example, at position 2028 a G is found unambiguously in the direct sequence of both the 16mer and 13mer from hybrid 2 (Fig. 4), whereas in the previously described cloned copy of the 16mer an A was at this position (11). Interestingly, a heterogeneous A/G is seen at this position in the repPCR sequence of the 16mers from hybrid 1 (data not shown). At position 2012, a heterogeneous G/C is seen in the 16mers from hybrid 2 with a C in the 13mer (Fig. 4), whereas in the cloned 16mer (and the 16mers from hybrid 1) a G was seen at this position (11). A single change was seen in the 13mers from hybrid 2 (relative to the single cloned copy) which consisted of an insertion of an A (with no heterogeneity) corresponding to position 2076. Each

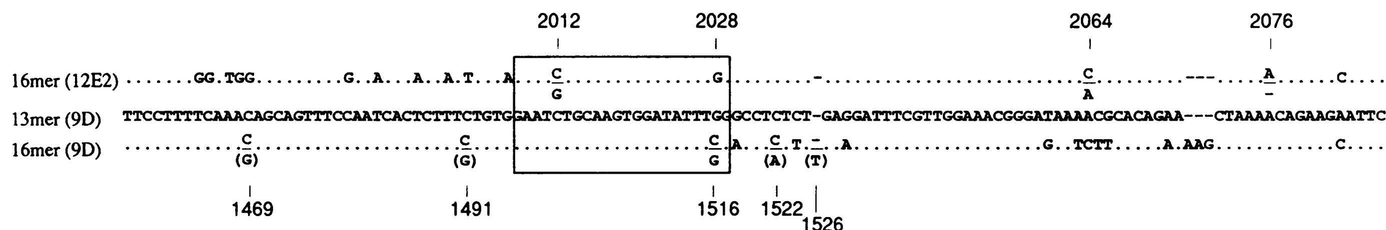

Figure 4. RepPCR Analysis of a Recombination Boundary. The direct sequence of the 13mers from hybrid 2 is shown, and the corresponding regions of the 16mers from hybrid 2 in both alignments is shown as comparison. Primer 17-9D was used to sequence the 16mer and 13mer; primer 17-12E2 was used only to sequence the 16mer (see Fig. 1A). A heterogeneous position is indicated by a double nucleotide ($^{a}/_{b}$). Parentheses indicate a nucleotide that was less abundant (but clearly visible) at a heterogeneous position, a dash indicates a one base pair deletion (relative to the other clones), a dot indicates an identical base pair as in the 13mer sequence. The location of the recombination junction was revised to a window of 22 bp (black box). See text for details.

of these base positions were confirmed by analysis of at least 18 independent clones derived from the PCR reactions (six clones each of the 16mer from each hybrid and of the 13mer; data not shown). In general, some differences previously described between the single cloned copies of the 16mer and 13mer (11) are either replaced by heterogeneous positions or eliminated in the repPCR sequence. The 13mers are more similar to the 16mers from the same chromosome (hybrid 2) than from a different chromosome (hybrid 1). Thus, it can be concluded that a subset of 16mers is more similar to the 13mers than the sequence of single cloned copies of each previously suggested. These discrepancies are due to either interchromosomal (homologue) differences or low copy mutations in the previously sequenced single clones.

The representative sequence of the 16mer and 13mer revealed by direct sequencing of repPCR products confirms the previous conclusions (11) about the register of misalignment between the 16mers that give rise to the 13mer. Some of the sequence differences between the two cloned repeat units were used, however, in the determination of the precise location of the junction boundaries of the recombination event, as discussed previously (11). Heterogeneous positions in the 16mer, where one nucleotide matches the nucleotide in the 13mer, can no longer be considered mismatches. As such, due to the replacement of sequence mismatches previously reported at positions 2028 and 2012 with heterogeneous positions, the junction boundary must now be expanded at the left side (as written, Fig. 4) to a length of 22bp (to position 2008). However, the right boundary remains unchanged, at position 1517. These results show that the 13mer seen from the repPCR amplification has the same molecular basis as previously reported in the independent cloning experiment. However, repPCR allows a more in depth and robust analysis of the junction boundary due to the ability to examine multiple copies of the repeat units simultaneously, removing possible interpretation errors due to examination of (non-representative) single clones or interchromosomal (homologue) variation.

Determination of the molecular basis of an X chromosome alpha satellite crossover product

Having demonstrated the validity of direct sequence data and examination of unequal crossing-over products obtained by repPCR, we applied the technique to the analysis of a repeat unit length variant seen in X chromosome alpha satellite DNA. The major form of alpha satellite DNA repeat unit from the X chromosome consists of a 12 monomer 2.0 kb unit, a single clone of which has been sequenced in its entirety (7). A minor form, which is seen in all X chromosomes examined (7, 13, 23), consists of a 17 monomer 3.0 kb repeat unit, presumably a result of unequal crossing-over between 2.0 kb repeat units. Both of these forms are readily visible on a Southern blot of human genomic DNA digested with Bam HI or Pst I and hybridized with an X chromosome alpha satellite repeat (23, 24). PCR amplification of human genomic DNA with primers X-5A and X-5B results in two major bands, a 1700 bp band which corresponds to the 2.0 kb repeat, and a 570 bp band, analyzed here, that corresponds to a 3.0 kb repeat unit (Fig. 5A,B). These bands were also both amplified from mouse/human somatic cell hybrids that contain one or more human X chromosomes as their only human complement.

The fact that the 570 bp band corresponds specifically to the 3.0 kb repeat unit was supported in several ways. Firstly, this band, although smaller than the 1700 bp band corresponding to the 2.0 kb repeat unit, is the size predicted from the placement of primers X-5A and X-5B if two 2.0 kb repeat units misalign by five monomers to form a 17 monomer 3.0 kb repeat unit (Fig. 5B). Secondly, DNA from a mouse/human somatic cell hybrid containing up to five human X chromosomes (per cell) was digested with Bam HI and with Pst I, and the 2.0 kb and 3.0 kb fractions from each were gel purified. PCR of these fragments with primers X-5A and X-5B resulted in greatly enriched amplification of the 570 bp band (>10 fold) in the 3.0 kb DNA fraction over the 2.0 kb fraction (data not shown). Finally, in order to rule out the artifactual generation of the 570 bp band during the PCR reaction, two lambda clones (λ 72 and λ 25) that contain only tandem 2.0 kb repeat units were PCR amplified with primers X-5A and X-5B. Only the 1700 bp band corresponding to the 2.0 kb repeat units was detected (Fig. 5A).

DNA from three mouse/human somatic cell hybrids containing human X chromosomes was PCR amplified with primers X-5A and X-5B, and the bands corresponding to the 2.0 kb (1700 bp PCR product) and the 3.0 kb (570 bp PCR product) repeat units were gel purified and subjected to direct sequencing. The 3.0 kb repeat unit and the corresponding regions from the 2.0 kb repeat unit were sequenced with primers X-5A and X-5B (Fig. 5B). By repPCR, the 3.0 kb repeat unit contained five heterogeneous positions within the region sequenced, while the 2.0 kb repeat unit contained four heterogeneous positions. There were no differences in these regions detected between the different X chromosomes examined, including the presence of the heterogeneous positions.

The sequence of the 3.0 kb repeat unit was aligned with the sequence of the cloned 2.0 kb repeat unit by using the positions

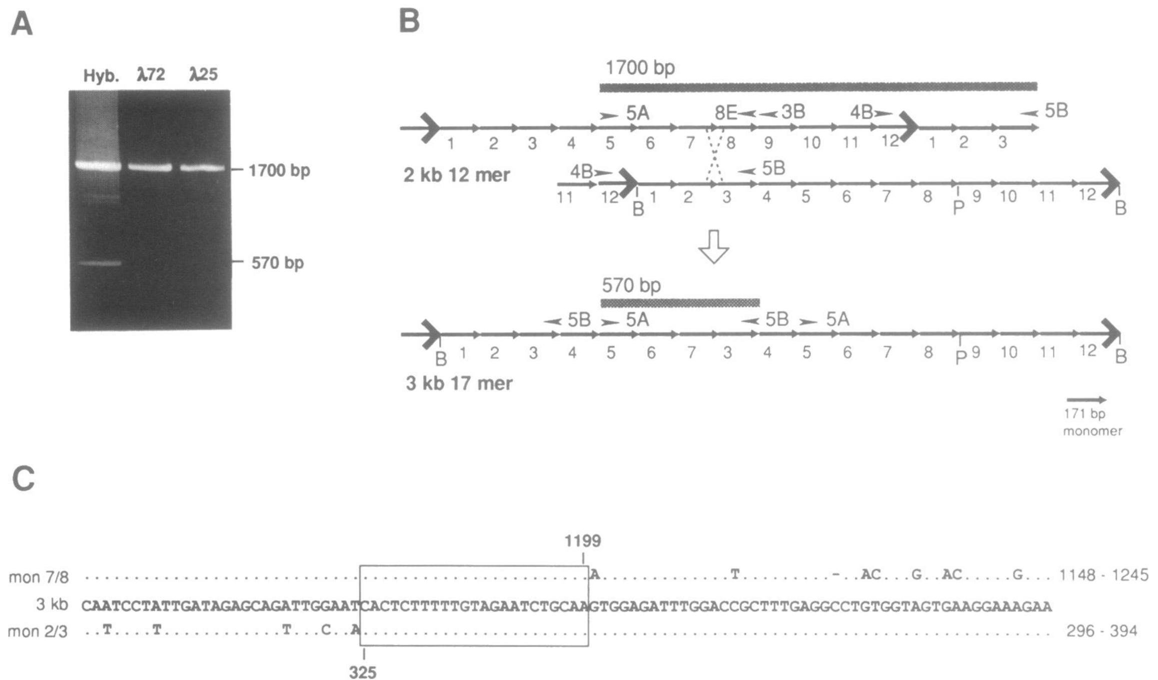

Figure 5. repPCR Analysis of an Unequal Crossing-Over Product from X Chromosome Alpha Satellite. **A**) Ethidium bromide stained agarose gel of PCR amplified X chromosome alpha satellite DNA using PCR primers X-5A and X-5B. The hybrid (Hyb.) is a somatic cell hybrid containing a single X chromosome, which shows amplification of both the 1700 bp band (corresponding to the 2.0 kb repeat unit) and the 570 kb band (corresponding to the 3.0 kb repeat unit). λ 72 and λ 25 are two lambda clones that contain only tandem 2.0 kb repeats, which only show amplification of the 1700 bp band. **B**) Misalignment and crossing-over of two 2.0 kb repeat units to give variant 3.0 kb repeat unit. Location of primers in the repeat units is as shown. Due to the placement of primers X-5A and X-5E, the 3.0 kb repeat unit gives a shorter PCR product than the 2.0 kb repeat unit. **C**) Determination of the recombination boundary between two misaligned repeat units. By comparison of the sequence of the 3.0 kb repeat unit variant to the 2.0 kb repeat unit, the precise location of the recombination can be narrowed to a 23 bp window (black box). The Genbank accession number of the 3.0 kb repeat unit is M99374.

of either primer X-5A or X-5B. When the sequences were aligned on primer X-5A, there was >99% identity for 370 bp between the two sequences from monomer 5 until the beginning of monomer 8, where identity dropped to 77%. Conversely, when the sequences were aligned at primer X-5B, there was >99% identity from monomer 3 to the end of monomer 2 (84 bp), where identity fell below 80%. Thus, a putative unequal crossing-over event appears to have occurred between the ends of monomers 2 and 7 or beginning of monomers 3 and 8 (Fig. 5B,C).

Additional sequence was obtained from the 1700 bp PCR product, using primers X-8E and X-3B for sequencing (see Fig. 5B), in order to obtain a repPCR based sequence of the area of recombination. This was to avoid any errors in the location of the junction boundary due to the possibility of non-representative base changes in the single cloned copy of the 2.0 kb repeat unit (7). (In fact, no heterogeneous positions were seen around the junction boundary). Examination of the homology between the two repeat types revealed an abrupt and symmetric breakpoint occurring between positions 325 and 1199 (numbering as in the 2.0 kb sequence in (7)) (Fig. 4C), allowing assignment of the recombination to a 23 bp window.

DISCUSSION

RepPCR: rapid assessment of sequence variation within repetitive DNA families

This report has demonstrated the further application of the PCR reaction to the study of tandemly repeated DNA families. Previous studies of the characterization of these DNA families have relied on analysis of a limited number of repeat units, obtained by cloning from lambda, cosmid, or plasmid libraries (11, 12, 18, 25, 26). Such analysis can introduce inaccuracy in the description of these repeat families due to the non-representative nature of individual clones; to obtain a representative sample of a large tandem DNA family would require analysis of an unwieldy number of individual clones (e.g. (18)). However, with the application of repPCR, multiple copies of the repeat units can be amplified and analyzed simultaneously by direct sequencing of the PCR product. Additionally, multiple independent copies of repeat units can be cloned easily from PCR products from many different sources. This technique allows a representative sampling of the repeat units in a repetitive DNA family to be analyzed with relatively little effort. Many DNA families within complex genomes could be studied with this approach, including ribosomal gene clusters, homologous multigene families, or even interspersed families such as LINE sequences. In general, knowledge of the sequence of one or a few repeat units allows design of specific PCR primers, which can be assessed for biases and refined by further analysis of the PCR products (see Fig. 2C). Additionally, use of such primers to amplify DNA from somatic cell hybrids would allow assessment of inter-chromosomal variation within these DNA families.

The validity of this approach was tested by using chromosome-specific PCR primer pairs for alpha satellite DNA from chromosomes 17 and X (14). After gel purification and direct sequencing of the amplified alpha satellite DNA, heterogeneous nucleotide positions between repeat units were seen as two bands

in the same base position on a sequencing ladder. The presence of these heterogeneous positions was confirmed by analysis of multiple independent clones obtained both by PCR (Tables 1 and 2) and non-PCR methods (18). As well, the relative intensities of the heterogeneous bases in the sequencing ladder appear to reflect their relative abundance in the repeat units, as judged both by the distribution in cloned copies (Table 1) and in a control amplification experiment (Fig. 2B). Such estimates, however, remain somewhat qualitative and require further evaluation of their usefulness. In general, repPCR appears to be able to determine heterogeneous positions between repeat units when they are present in at least 10–20% of the repeat units. This allows for a representative analysis of the repeat units, despite the approximately 1% divergence overall between most repeat units (e.g. (18)). Thus, sequence obtained by this method, including heterogeneous positions, will approximate an overall consensus of repeat units, but not include any low copy mutations that may be present in individual or a low proportion of clones.

PCR primers that are targeted to multicopy DNA could introduce a bias if they anneal to a limited subset of repeat units, due to a heterogeneous position in the repeat unit within the primer site. Primers 17-1A and 17-2A, which amplify chromosome 17 alpha satellite (14), were shown to be unbiased by PCR amplifying and sequencing directly their annealing sites from genomic DNA. There is a formal possibility, however, that this test could be misleading if linkage disequilibrium exists in repeat units between a heterogeneous position in the original primer sites and a heterogeneous position in the primer sites used to test the original primers. Historically, direct sequencing of double stranded PCR products has been troublesome, generally attributed to the reannealing of the DNA strands and low quality of gel purified DNA. Several improvements have been reported to overcome this technical difficulty (reviewed in (27)). In this study, the addition of TdT was found to greatly alleviate the problem of polymerase pauses and stops, which is clearly a necessity in order to recognize heterogeneous positions correctly.

The ability to obtain a representative sample of the repeat units in a tandem DNA family greatly facilitates the study of variation within that DNA family. By examination of the heterogeneous positions seen by direct sequencing of PCR amplified DNA, the intrachromosomal variation can be rapidly assessed. As well, by examination of multiple chromosomes, both from somatic cell hybrids and in populations, estimates of interchromosomal (homologue) variation can be obtained. As such, the variation seen in this initial study between the 16mers found in different chromosomes 17 from hybrids 1 and 2 suggests that there may be significant differences between repeat units found on different copies of homologous chromosomes. Interestingly, since hybrids 1 and 2 represent the two major haplotypes of chromosome 17 alpha satellite (19), such variation may be a reflection of haplotype specific changes (see (9)). RepPCR analysis of several more chromosomes 17 from each haplotype would be required to determine the extent of this haplotype-specific variation. At least three X chromosomes were also examined using PCR sequencing, and no variation was seen between these chromosomes, extending previous studies in which conclusions were based on multiple individual clones (18). Again, alpha satellite from multiple additional X chromosomes will need to be analyzed by repPCR to establish the extent of interchromosomal variation. RepPCR provides an opportunity for extensive analysis of both the intra- and interchromosomal variation in tandemly repeated DNA families.

Application of repPCR to the analysis of unequal crossing-over

PCR amplification of tandemly repeated DNA also can reveal variation in the length of repeat units if primers span the putative site of recombination. For example, primers X-5A and X-5B revealed two repeat units from chromosome X alpha satellite, which allowed molecular analysis of a previously uncharacterized 3.0 kb repeat unit variant. Due to the placement of the primers within the repeat unit, the longer repeat unit actually corresponded to a shorter PCR product, which may have facilitated its amplification. It should be pointed out that no independent evidence exists to suggest that this is the only form of 3.0 kb repeat unit seen on Southern blots of chromosome X alpha satellite (23, 24); it might be possible to design additional primer sets spanning other potential recombination junctions. This is especially true when the length variant is approaching sizes that become difficult to efficiently PCR amplify.

Primers 17-1A and 17-2A amplify multiple repeat types from chromosome 17 alpha satellite that correspond to the different length repeat types representative of the different haplotypes (14). Here we characterized the PCR product that corresponds to the 13mer, and compare the results to a previously described single cloned copy of this variant (11). The proposed register of misalignment and location of the crossover were the same as previously described, demonstrating the validity of repPCR for this application. However, the precise recombination boundaries had to be adjusted, due to the refinement of the nucleotide sequence used to elucidate the recombination boundaries. In this study, the 16mer and 13mer from the same chromosome 17 were compared, thus eliminating any potential interchromosomal sequences differences. As well, the elucidation of heterogeneous positions by repPCR required a more conservative treatment of the data, in that sequence identity between the variant repeat units to one of the heterogeneous nucleotides could no longer be considered a mismatch, since a subset of repeat units will exist that match at this position; position 2012 is treated in this manner (Fig. 4). As such, delineation of recombination junctions using this technique will be more robust, but may result in expansion of the size of the 'window' within which the recombination junction can be localized. Significantly, the location of the recombination junctions for both unequal crossing-over products overlap (10, 11), adding to a growing set of over 12 recombination windows that occur within this location in the 171 bp alpha satellite monomer (P.E.W, J.S.Waye, & H.F.W., unpublished data). The ability to PCR amplify repeat unit length variants from tandemly repeated DNA families should allow an in depth molecular study of the unequal crossing-over events that are thought to play important roles in the evolution of tandemly repeated DNA families (5, 28–30).

Tandemly repeated DNA families are found in the genomes of most complex organisms, often in functionally interesting chromosomal locations (1). In light of their large array sizes and abundant nature, as well as the elucidation of structural protein binding sequence motifs within the tandem repeat units (31), repPCR of tandemly repeated DNA might be a useful approach for addressing some of the structural requirements necessary for proper chromosomal function and/or segregation (4). The ability to rapidly and representatively examine repeat unit variation within and between these tandem DNA families should provide some insight into their possible roles in chromosome structure and function (32). Additionally, tandem DNA families undergo rapid evolutionary processes that involve the complex interplay

of several important genetic mechanisms such as unequal sister chromatid exchange and meiotic and mitotic recombination (3, 28–30). As such, the study of repeat unit variation within individual arrays and through populations should provide important information about the relative contribution of these mechanisms towards shaping our genome, as well as provide insight into their molecular nature.

ACKNOWLEDGEMENTS

The authors wish to thank the other members of the laboratory for helpful discussions and support, especially those who were also involved in sequencing PCR reactions. This work was supported by grant HG00107 from the National Institutes of Health.

REFERENCES

1. Miklos, G.L.G. (1985) In J. R. Macintyre (ed.), Molecular evolutionary genetics, Plenum Publishing Corp., New York, pp. 241–313.
2. Brutlag, D.L. (1980) *Annu. Rev. Genet.*, **14**, 121–144.
3. Dover, G.A. (1986) *Trends Genet.*, **2**, 159–165.
4. Willard, H.F. (1990) *Trends Genet.*, **6**, 410–416.
5. Willard, H.F. and Waye, J.S. (1987) *Trends Genet.*, **3**, 192–198.
6. Willard, H.F. (1985) *Am. J. Hum. Genet.*, **37**, 524–532.
7. Waye, J.S. and Willard, H.F. (1985) *Nucl. Acids Res.*, **13**, 2731–2743.
8. Mahtani, M.M. and Willard, H.F. (1990) *Genomics*, **7**, 607–613.
9. Warburton, P.E. and Willard, H.F. (1990) *J. Mol. Biol.*, **216**, 3–16.
10. Waye, J.S. and Willard, H.F. (1986) *Mol. Cell. Biol.*, **6**, 3156–3165.
11. Waye, J.S. and Willard, H.F. (1986) *Nucl. Acids Res.*, **14**, 6915–6927.
12. Tyler-Smith, C. and Brown, W.R.A. (1987) *J. Mol. Biol.*, **195**, 457–470.
13. Willard, H.F., Waye, J.S., Skolnick, M.H., Schwartz, C.E., Powers, V.E. and England, S.B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5611–5615.
14. Warburton, P.E., Greig, G.M., Haaf, T. and Willard, H.F. (1991) *Genomics*, **11**, 324–333.
15. Witt, M. and Erickson, R.P. (1989) *Hum. Genet.*, **82**, 271–274.
16. Dunham, I., Lengauer, C., Cremer, T. and Featherstone, T. (1992) *Hum. Genet.*, **88**, 457–462.
17. Engelke, D.R., Hoener, P.A. and Collins, F.S. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 544–548.
18. Durfy, S.J. and Willard, H.F. (1989) *Genomics*, **5**, 810–821.
19. Willard, H.F., Greig, G.M., Powers, V.E. and Waye, J.S. (1987) *Genomics*, **1**, 368–373.
20. Lafreniere, R.G., Brown, C.J., Powers, V.E., Carrel, L., Davies, K.E., Barker, D.F. and Willard, H.F. (1991) *Genomics*, **11**, 352–363.
21. Fawcett, T.W. and Bartlett, S.G. (1990) *Biotechniques*, **9**, 46–48..+†.
22. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis,K.B. and Erlich, H.A. (1988) *Science*, **239**, 487–491.
23. Yang, T.P., Hansen, S.K., Oishi, K.K., Ryder, O.A. and Hamkalo, B.A. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6593–6597.
24. Willard, H.F., Smith, K.D. and Sutherland, J. (1983) *Nucl. Acids Res.*, **11**, 2017–2033.
25. Choo, K.H., Brown, R., Webb, G., Craig, I.W. and Filby, R.G. (1987) *DNA J. Mol. Biol.l*, **6**, 297–305.
26. Jorgensen, A.L., Bostock, C.J. and Bak, A.L. (1986) *J. Mol. Biol.*, **187**, 185–196.
27. Bevan, I.S., Rapley, R. and Walker, M.R. (1992) *PCR Meth. Appl.*, **1**, 222–228.
28. Ohta, T. (1980) Evolution and Variation of Multigene Families. In S. Levin (series ed.) *Lecture Notes in Biomathematics*, Vol. 37. Springer-Verlag, New York.
29. Smith, G.P. (1976) *Science*, **191**, 528–535.
30. Willard, H.F. (1991) *Curr. Opin. Genet. Dev.*, **1**, 509–514.
31. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) *J. Cell Biol.*, **109**, 1963–1973.
32. Haaf, T., Warburton, P.E., and Willard, H.F. (1992) *Cell*, **70**, 681–696.