Development 101 Supplement, 77–92 (1987)
Printed in Great Britain © The Company of Biologists Limited 1987

Repeated DNA of the human Y chromosome

KIRBY D. SMITH, KEITH E. YOUNG, C. CONOVER TALBOT, Jr
and BARBARA J. SCHMECKPEPER

The Division of Medical Genetics, Department of Medicine, The Johns Hopkins University School of Medicine, Traylor Building, Room 933, Baltimore, Maryland 21205, USA

Summary

A significant fraction of the human Y chromosome is composed of DNA sequences which have homologues on the X chromosome or autosomes in humans and non-human primates. However, most human Y-chromosome sequences so far examined do not have homologues on the Y chromosomes of other primates. This observation suggests that a significant proportion of the human Y chromosome is composed of sequences that have acquired their Y-chromosome association since humans diverged from other primates.

More than 50 % of the human Y chromosome is composed of a variety of repeated DNAs which, with one known exception, can be distinguished from homologues elsewhere in the genome. These include the alphoid repeats, the major human SINE (*Alu* repeats) and several additional families of repeats which account for the majority of Y-chromosome repeated DNA. The alphoid sequences tandemly clustered near the centromere on the Y chromosome can be distinguished from those on other chromosomes by both sequence and repeat organization, while the majority of Y-chromosome *Alu* repeats have little homology with genomic consensus *Alu* sequences. In contrast, the Y-chromosome LINE repeats cannot be distinguished from LINEs found on other chromosomes. It has been proposed that both SINE and LINE repeats have been dispersed throughout the genome by mechanisms that involve RNA intermediates. The difference in the relationship of the Y-chromosome *Alu* and LINE repeats to their respective family members elsewhere in the genome makes it possible that their dispersal to the Y chromosome has occurred by different mechanisms or at different rates.

In addition to the SINE and LINE repeats, the human Y chromosome contains a group of repeated DNA elements originally identified as 3·4 and 2·1 kb fragments in *Hae*III digests of male genomic DNA. Although the 3·4 and 2·1 kb Y repeats do not cross-react, both exist as tandem clusters of alternating Y-specific and non-Y-specific sequences.

The 3·4 kb Y repeats contain at least three distinct sequences with autosomal homologies interspersed in various ways with a collection of several different Y-specific repeat sequences. Individual recombinant clones derived from isolated 3·4 kb *Hae*III Y fragments have been identified which do not cross-react. Thus, the 3·4 kb *Hae*III Y fragments are a heterogeneous mixture of sequences which have in common the regular occurrence of *Hae*III restriction sites at 3·4 kb intervals and an organization as tandem clusters at various sites along the Y-long arm.

The 2·1 kb *Hae*III Y fragment cross-reacts with a 1·9 kb *Hae*III autosomal fragment. Both the Y-chromosomal and autosomal fragments are part of tandem clusters which have a unit length of 2·4 kb. All of the 2·4 kb Y repeats are similar and contain a 1·6 kb Y-specific repeat and an 800 bp sequence which has homology with an 800 bp sequence in the autosomal 2·4 kb repeats. While this 800 bp sequence is common to both Y and autosomal 2·4 kb repeats and is associated with a single Y-specific repeat, it is associated with at least four non-cross-reacting autosome-specific sequences. Like the Y repeat, the autosomal repeats exist as tandem clusters of 2·4 kb units and are composed of an 800 bp common sequence alternating with a 1·6 kb autosome-specific sequence. Thus, in humans, the common sequence is associated with several different sequences yet always occurs as part of a tandem cluster of 2·4 kb repeats. The common and autosome-specific sequences of the 2·4 kb repeats are also present in gorillas as part of organized repeat units. However, in gorillas the two are not associated with each other.

The Y-chromosome repeats described here are a heterogeneous mixture of sequences organized into specific sets of alternating Y-specific and non-Y-specific sequences. They do not have an identified function and the mechanisms by which they are generated are unknown. Nevertheless, their marked chromosomal specificity and the regularity of the basic repeat unit in each type of repeat seem inconsistent with stochastic mechanisms of sequence diffusion between chromosomes.

Key words: Y chromosome, repeated DNA, evolution, primate.

Introduction

The evolution of eukaryotes has been accompanied by an enormous increase in genome size (Britten & Davidson, 1969). This increase is largely the result of an expansion of sequences that do not code for protein structure (Galau *et al.* 1976). While some of this 'non-coding' DNA undoubtedly regulates gene expression, it is likely that most of it has either some other, as yet unknown, function or no function. A significant fraction of such 'non-coding' DNA is composed of various families of repeated sequences interspersed with single copy DNA. The ubiquity of this interspersion pattern (Britten & Davidson, 1971) has suggested that it may play some fundamental role in the function of the genome, such as regulation of gene expression (Davidson & Britten, 1979) or in chromosome structure (Manuelidis & Ward, 1984). An alternative view, based on the way many repeats may multiply and spread within the genome, is that repeated DNA is parasitic (Orgel & Crick, 1980) or selfish (Doolittle & Sapienza, 1980). The human Y chromosome provides a setting for examining the organization, evolutionary dispersion and possible function of various categories of repeated DNAs for several reasons. First, it has a very high proportion of repeated sequences (Kunkel, Smith & Boyer, 1979) and only a limited number of known functional genes (Goodfellow, Darling & Wolfe, 1985). Second, it has a very limited region that participates in meiotic pairing (Cooke, Brown & Rappold, 1984; Simmler *et al.* 1985). Thus the majority of Y-chromosome sequences are not involved in the usual meiotic recombination events. Third, the human Y chromosome is of recent evolutionary origin in that most of its sequences have homology with the X chromosome or autosomes of other primates but *not* with their Y chromosomes (Szabo *et al.* 1980; Kunkel & Smith, 1982; Burk, Patrick & Smith, 1985*a*; Page, Harper, Love & Botstein, 1984; Singer, 1982).

In broad terms, there are three general categories of repeated DNA: short, interspersed elements, called SINEs (Singer, 1982); long interspersed elements called LINEs (Singer, 1982) and tandemly clustered repeated elements. In this paper, we will review the distribution and organization properties of such repeats on the human Y chromosome.

SINEs

The major human SINE, the *Alu* family (Schmid & Jelinik, 1982), is repeated about 300 000 times and dispersed throughout the genome (Rinehart, Ritch, Deininger & Schmid, 1981). It is homologous to an abundant small nuclear RNA (Jelinek *et al.* 1980) and to double-stranded heterogeneous nuclear RNA (Jelinek *et al.* 1980; Fritsch, Lawn & Maniatis, 1980) and is transcribed *in vitro* by RNA polymerase III (Duncan *et al.* 1980). These findings, together with the widespread dispersion of SINEs and the presence of short direct repeats flanking some SINEs (Singer, 1982), suggest that *Alu* repeats may have been dispersed about the genome *via* self-primed RNA polymerase III transcription followed by reverse transcription and duplication of a target site (Jagadeeswaram, Foget & Weissman, 1981; Van Arsdell *et al.* 1981). It has also been suggested that *Alu* repeats may have initially arisen as pseudogenes of 7SL RNA (Ullu & Tschudi, 1984; Ullu & Weiner, 1985) or transfer RNA (Daniels & Deininger, 1983).

Although detailed studies have not yet been completed, it is clear that the spectrum of *Alu* sequences on the human Y chromosome differs from that of the rest of the genome. We have carried out preliminary studies of the *Alu* repeats on the human Y chromosome by using cloned *Alu* sequences to probe restriction digests of genomic DNA from a somatic cell hybrid in which the only cytologically detectable human chromosome is the Y (Burk *et al.* 1985*a*). *Alu* probes which reflect average *Alu* sequences, such as Blur 2 and 8 (Deininger *et al.* 1981), detect few homologous sequences in these digests when hybridized under conditions that detect the majority of *Alu* repeats in human genomic DNA. Probing under more relaxed conditions yields a significant increase in the number of hybridizing Y-chromosome sequences. Similar results have been obtained by probing recombinant DNA libraries constructed from somatic cell hybrids in which the only human chromosome is the Y (Burk *et al.* 1985*a*; Wolfe, Erickson,

Rigby & Goodfellow, 1984) with 'typical' *Alu* sequences. Thus, while the human Y chromosome contains a significant number of *Alu* sequences, they are considerably diverged relative to the average genomic *Alu* sequence.

LINEs

The human LINE (Adams *et al.* 1980), also designated the *Kpn* family (Shafit-Zagardo, Maio & Brown, 1982) or L1 family (Voliva *et al.* 1983), is abundant [estimates range from several thousand (Adams *et al.* 1980; Manuelidis & Biro, 1982) to 100 000 (Hwu, Roberts, Davidson & Britten, 1986)], widely dispersed in the genome (Shafit-Zagardo *et al.* 1982), and usually flanked by single-copy DNA (Adams *et al.* 1980). Many LINE sequences are about 6 kb (Adams *et al.* 1980) and colinear (Miyaki, Migita & Sakaki, 1983). However, 'scrambled' LINE elements (Potter, 1984) and shorter elements with truncated 5′ ends also exist (Grimaldi & Singer, 1983). We have shown (Schmeckpeper, Willard & Smith, 1981) that the major LINE restriction fragments in the human genome do not have a marked chromosome-specific organization, but one example of chromosome specificity in LINE organization has been reported in the African Green Monkey (Lee & Singer, 1986).

We (Schmeckpeper, Scott & Smith, 1984) and others (Kole, Haynes & Jelinek, 1983; Shafit-Zagardo, Brown, Zavodny & Maio, 1983) have detected primate LINE transcripts, the vast majority of which lack a poly(A) tail, are localized to the nucleus and are synthesized by RNA polymerase II. LINEs are often flanked by short direct repeats, and some have a poly(A)-rich sequence at their 3′ end (Grimaldi, Skowronski & Singer, 1984). These observations, coupled with the 5′ truncation of many LINEs, are consistent with their insertion into new chromosomal locations by a reverse-transcription mechanism similar to that thought to produce pseudogenes (Kole *et al.* 1983). However, not all LINEs have a structure compatible with this mechanism of insertion. Some lack flanking direct repeats (Grimaldi *et al.* 1984), some lack poly(A) tracts and several have permuted arrangements (Kole *et al.* 1983). Based on sequence comparisons among various LINEs it has been suggested that the family as a whole may be evolving in a manner termed 'concerted evolution' (Martin *et al.* 1985). Mechanisms proposed for 'concerted evolution' include gene conversion, duplicative transposition and unequal sister chromatid exchange (Dover & Flavell, 1984), mechanisms which could also contribute to the dispersal of LINEs within the genome.

We have previously shown that the major LINE restriction fragments of the human X chromosome and a selected subset of autosomes are identical to those detected in whole genomic DNA (Schmeckpeper *et al.* 1981). Detailed restriction analyses of genomic DNA from the somatic cell hybrid in which the only cytologically detectable human chromosome is the Y demonstrate that the LINE repeats on the Y are indistinguishable from those found elsewhere in the genome. These studies have utilized cloned LINE probes from the 3′ and 5′ halves of a complete LINE and restriction digests that reveal specific LINE organizations (Schmeckpeper *et al.* 1981). An example of these results is shown in Fig. 1 which shows there is no detectable difference in the restriction fragments from the X and Y chromosomes

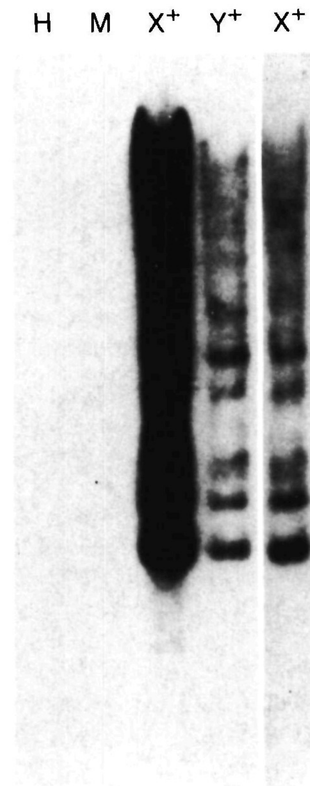

Fig. 1. Analysis of LINEs in the human Y chromosome. Genomic DNA from somatic cell hybrids in which the only human chromosome is the X (X^+) or Y (Y^+) and from the parental rodent cell lines, hamster (H) or mouse (M) was digested with restriction enzyme *KpnI*. The resulting DNA fragments were size fractionated by electrophoresis on a 1 % agarose gel, transferred to nitrocellulose filters and probed with a mixture of radiolabelled LINE probes from the 3′ and 5′ ends of a LINE. Each lane contains 6·5 μg of genomic DNA. The X^+ lane on the far right was exposed for one-sixth the time of the other four lanes. The relative hybridization intensity of the X^+ and Y^+ DNAs indicate that the amount of LINE sequence on the Y chromosome is proportional to its size.

detected with a mixture of the 3′ and 5′ probes. Since we have previously demonstrated that the X-chromosome LINEs cannot be distinguished from the LINEs detected in whole genomic DNA (Schmeckpeper *et al.* 1981), the Y-chromosome LINEs are also representative of genomic LINEs. Knowing the abundance of LINEs on the X chromosome (Schmeckpeper *et al.* 1981) and comparing the relative hybridization intensities obtained in Fig. 1, it can be deduced that the abundance of LINEs on the Y is about that predicted from the size of the Y chromosome and the abundance of LINEs in the genome as a whole.

Alphoid repeats

Tandemly clustered repeats, by still poorly defined mechanisms, can change their copy number at a particular locus, change their chromosomal location quickly over evolutionary time and acquire chromosome-specific characteristics which are somehow distributed among most family members at a particular locus. Most mammalian species have a family of homologous tandemly repeated DNAs (Brutlag, 1980) clustered at centromeres and other heterochromatic sites (Miklos & John, 1979). Extensive restriction analyses reveal subsets of sequences (termed domains) which can be chromosome specific (Brown & Dover, 1980; Lee & Singer, 1982). In primates, such a repeated DNA family, alphoid DNA (Singer, 1982), exists as tandem arrays of a small unit often found near centromeres. The units within a single array are similar to each other, perhaps being maintained by unequal crossing-over (Manuelidis, 1982). These alphoid repeats have been shown to have chromosome-specific organizations (Willard, Smith & Sutherland, 1983; Wolfe *et al.* 1985) and the sequences seem to be evolving rapidly since many differences occur between species. As shown by Wolfe *et al.* (1985), the Y chromosome has a unique set of alphoid repeats which can be distinguished from the alphoid repeats of other chromosomes by sequence and by the periodicity of particular restriction sites within the tandem array Since the range of sequence divergence among Y-chromosome alphoid monomers is similar to that found on other chromosomes (Wolfe *et al.* 1985) and since most chromosomes have a specific set of alphoid repeats which are distinguishable by sequence and restriction site periodicity, the Y chromosomes is not unique with respect to the characteristics of its alphoid repeats.

Thus, for the major categories of repeat sequences which the Y chromosome shares with the rest of the genome, the characteristics of LINE and alphoid sequences are typical of these found elsewhere in the genome. In contrast, the predominant *Alu* sequences on the Y are considerably diverged relative to the average genomic *Alu* sequence. Thus, if both LINEs and *Alu* repeats became associated with the Y chromosome by a mechanism involving RNA intermediates, they must have invaded the Y at different times or at quite different rates. Alternatively, they may have been incorporated into the Y chromosome by different mechanisms or been subject to different selective pressures.

Y-chromosome-specific repeats

In addition to the repeats described above, the Y chromosome contains two major sets of repeats that are specific to the Y chromosome (Cooke, 1976). As seen in Fig. 2, digestion of male genomic DNA with restriction enzyme *Hae*III yields two distinct ethidium-bromide-staining bands (3·4 and 2·1 kb) that are not seen in similar digests of female DNA. Since both *Hae*III fragments are male specific (Fig. 2), have a concentration within the genome proportional to the number of Y chromosomes present and are present in a somatic cell hybrid where the only human chromosome is the Y (data not shown), they must be derived from the Y chromosome. Fig. 2 also demonstrates that the 3·4 and 2·1 kb *Hae*III fragments, isolated from genomic DNA by a combination of physical techniques (Kunkel & Smith, 1982), do not cross-hybridize.

While the 3·4 and 2·1 kb fragments are themselves from the Y chromosome, each contains sequences which cross-react with DNA from other chromosomes. Failure to hybridize with DNA from a mouse × human somatic cell hybrid in which the only human chromosome is the X (Dorman, Shimizu & Ruddle, 1978) established that the non-Y-homologues of both the 3·4 and 2·1 kb *Hae*III Y fragments are autosomal (data not shown). We have also shown (Kunkel & Smith, 1982) that the 3·4 kb *Hae*III fragment is composed of Y-specific sequences interspersed with non-Y-specific ones and that the major sites of autosomal homology are chromosomes 1, 9 and 15 (Burk *et al.* 1985*b*). It is evident in Fig. 2 that the 2·1 kb *Hae*III fragment, isolated directly from genomic DNA, also detects homologues in *Hae*III digests of female DNA. To determine if the hybridization to female DNA resulted from homology to the same 2·1 kb *Hae*III fragment detecting the male-specific band, *Hae*III digests of male and female DNA were probed with a 2·1 kb *Hae*III fragment cloned into pBR322. As seen in Fig. 3, the cloned probe hybridized to the male-specific 2·1 kb band and to a 1·9 kb band in both male and female DNA. Thus, like the 3·4 kb *Hae*III fragment, the 2·1 kb *Hae*III fragment, though itself derived from the Y, contains sequences with homology to other chromosomes.

The major concentration of the 1·9 kb autosomal homologues of the Y-specific 2·1 kb *Hae*III fragment was localized to chromosome 14 by hybridization to restriction digests of DNA from a panel of somatic cell hybrids segregating different combinations of human chromosomes (O'Brien *et al.* 1983).

The distribution of the 3·4 and 2·1 kb *Hae*III fragments along the Y chromosome was determined by *in situ* hybridization (Burk *et al.* 1985*b*). Fig. 4 summarizes these results and shows that both fragments are distributed throughout the length of the Y long arm. These results have been confirmed by reassociation kinetics and Southern blot analysis of DNA from individuals with various Y-chromosome deletions.

Cooke (1976) previously presented evidence that the 3·4 kb *Hae*III fragment was, at least in part, tandemly arranged within the Y chromosome. This conclusion was based on the generation of multimers of this fragment in partial *Hae*III digests of genomic male DNA. To determine if the fragments homologous to the 2·1 kb *Hae*III fragment were also arranged in tandem, we performed a series of partial *Hae*III digests of male and female DNA and hybridized the resulting Southern blots with a 2·1 kb *Hae*III probe. The results obtained with male DNA (Fig. 5) show a ladder of fragments expected for a tandem repeat. However, the multimers are not based on a unit length of 2·1 kb; no band is observed at 4·2 kb

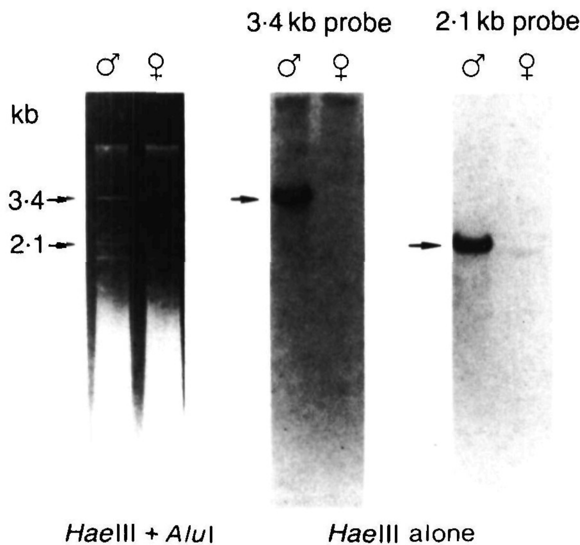

Fig. 2. Human *Hae*III restriction fragments. 5 μg of male or female DNA was digested with restriction enzymes *Hae*III and *Alu*I. In ethidium-bromide-stained gels (left panel) two male-specific fragments are seen (3·4 and 2·1 kb). *Alu*I was included to improve the visualization of these fragments, since it does not digest the 3·4 kb fragment and removes only 40 bp from the 2·1 kb fragment. Fragment lengths were estimated by comparison to DNA standards of known length (*Hin*dIII digests of Lambda DNA and *Hae*III digests of phiX174 DNA). *Hae*III restriction digests of male and female DNA were transferred to nitrocellulose filters and hybridized with 3·4 and 2·1 kb fragments purified from male genomic DNA. These fragments were isolated from *Hae*III-digested male DNA which was passed over a RPC-5 column and fractionated according to length. In the case of the 2·1 kb fragment, fractionation is strongly influenced by its base composition. To remove contaminating DNA, fractions containing the relevant fragments were pooled and digested with restriction enzymes that do not digest the Y fragments. The Y fragments were then recovered after size fractionation on agarose gels. The gel-purified fragments were radiolabelled by nick translation to a specific activity of 10^8 cts min^{-1} μg^{-1}. Before hybridization, filters were soaked in 3 × SSC, 5× Denhardts' mixture, 1 % glycine, 50 mM-phosphate buffer, pH 7·2 and 50 % formamide at 37°C for 12–18 h. Filters were hybridized in the same mixture with the radiolabelled probe at 5×10^5–10^6 cts min^{-1} ml^{-1} for 2–3 days at 37°C. After hybridization, the filters were washed extensively and exposed to X-ray film.

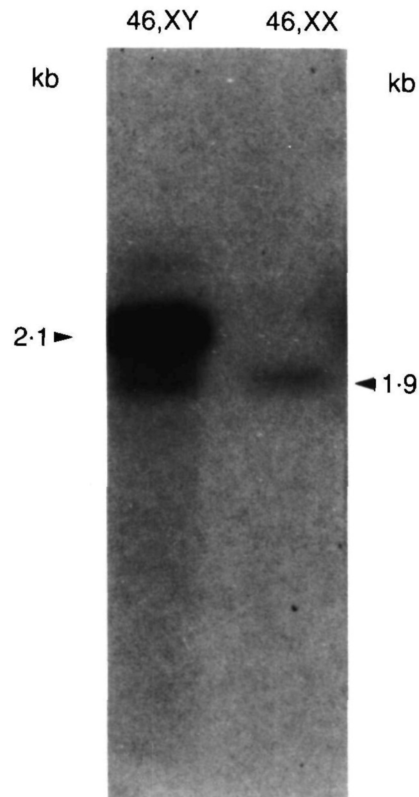

Fig. 3. Homology of human genomic DNA with cloned 2·1 kb *Hae*III Y DNA. Male (46,XY) or female (46,XX) DNA was digested with *Hae*III and size fractionated by electrophoresis on a 0·8 %, 30 cm agarose gel. After transfer to nitrocellulose filters, the digests were hybridized with a cloned 2·1 kb *Hae*III fragment derived from a preparation purified as described in Fig. 2 legend and then cloned into pBR322. The filter was hybridized with radiolabelled probe as described in Fig. 2.

corresponding to a dimer. Rather, prominent bands are seen at 2·4, 4·5, 4·8 and 7·2 kb, suggesting that the basic fragment length is 2·4 kb. Similar results were obtained with female DNA (Fig. 6). Like the Y fragments, the autosomal fragments appear to be tandem clusters of a 2·4 kb repeat. This interpretation was confirmed by comparing the length of homologous fragments in male and female DNA digested with various restriction enzymes and probed with a cloned 2·1 kb *Hae*III Y fragment (Fig. 7). With the exception of *Hae*III, the restriction enzymes indicated in Fig. 7 generated a fragment of 2·4 kb in both male and female DNA. Thus, the restriction pattern of Y and autosomal 2·1 kb *Hae*III homologues are

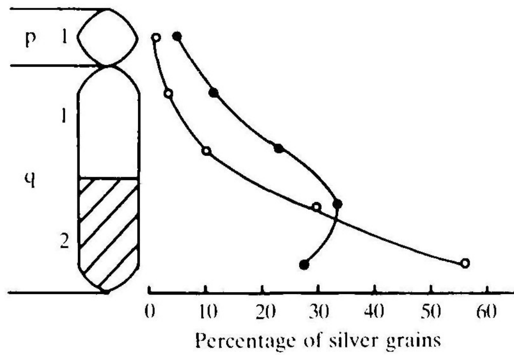

Fig. 4. Distribution of 3·4 and 2·1 kb *Hae*III Y DNA on the human Y chromosome. Distributions of Y DNA repeat fragments along the Y chromosome were determined by *in situ* hybridization. The position of the silver grains along the Y with 3·4 (●) or 2·1 (○) kb Y DNA was determined by scoring 64 and 50 Y chromosomes, respectively. Both fragments are distributed throughout the length of the long arm; in addition, the 2·1 kb fragment shows a significant concentration at the end of the fluorescent segment of the long arm (represented by the hatched distal half). These *in situ* results have been confirmed by Cot and Southern blot analyses of individuals with varying Y long arm deletions.

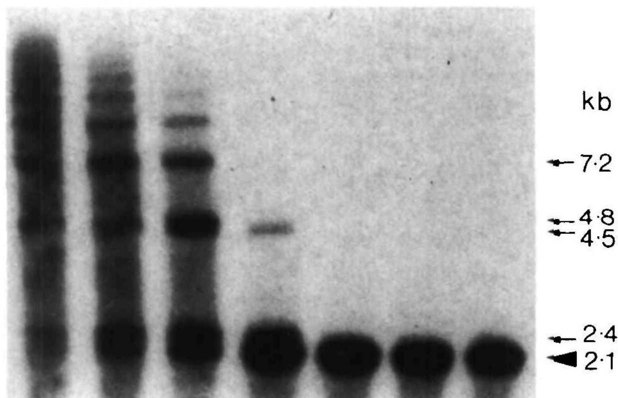

Fig. 5. Partial *Hae*III digestion of male genomic DNA. Male genomic DNA (5 μg in each lane) was digested with increasing amounts of *Hae*III. The lowest enzyme:DNA ratio is on the left and the highest ratio is on the right. The digests were size fractionated by electrophoresis through a 0·8 % agarose gel, transferred to a nitrocellulose filter and hybridized with a radiolabelled genomic 2·1 kb *Hae*III Y DNA probe as described in Fig. 2 legend. The indicated fragment lengths were determined by comparison to DNA standards of known length.

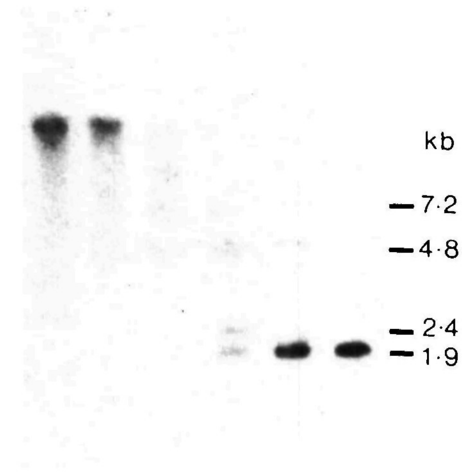

Fig. 6. Partial *Hae*III digestion of female genomic DNA. Female genomic DNA (5 μg in each lane) was digested and hybridized with a 2·1 kb *Hae*III Y DNA probe as described in Fig. 5 legend.

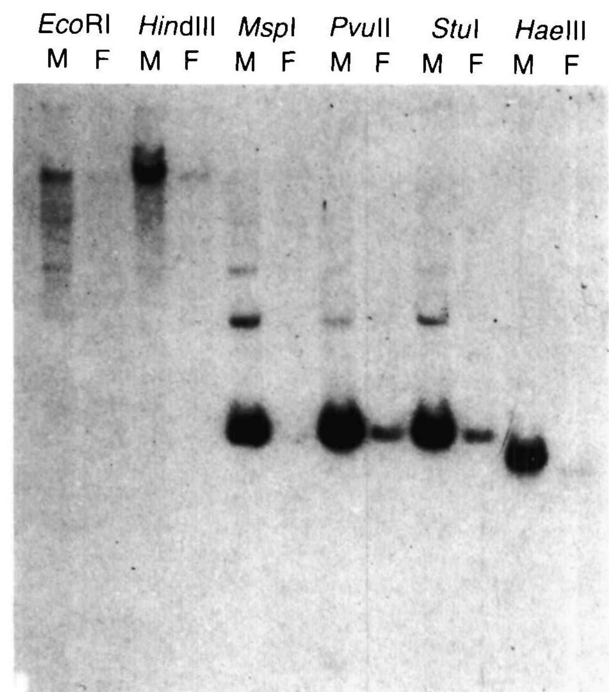

Fig. 7. Comparison of autosomal 2·4 kb Y homologues. Male (M) or female (F) DNA was digested with the restriction enzyme indicated at the top of the figure. Each lane contains 5 μg of DNA and was hybridized with a genomic 2·1 kb Y probe as indicated in Fig. 2 legend.

quite similar. This observation raises the possibility that the autosomal and Y-chromosome 2·4 kb repeats are essentially identical, differing only in a few restriction sites. That this is not the case can be demonstrated by probing restriction digests of a number of independently derived 2·4 kb Y-repeat clones with radiolabelled male or female genomic DNA. The 2·4 kb repeats on the Y chromosome, which contain the 2·1 kb *Hae*III fragment, have a single *Msp*I site which is not present in the autosomal 2·4 kb repeat (see Figure 13; Young, Willard & Smith, 1983). Thus, cloned 2·4 kb repeats derived exclusively from the Y chromosome can be obtained from male genomic DNA after digestion with *Msp*I. If restriction digests of these cloned Y repeats do not have the same hybridization pattern when probed with male and female DNA, then there must be sequence differences between the Y and autosomal 2·4 kb repeats. When restriction digests of the Y-specific 2·4 kb *Msp*I clones were probed with male and female DNA, a set of fragments of about 300 bp was hybridized by both probes, but larger fragments were only detected with the male probe (Fig. 8). Therefore, it is only these smaller Y-chromosome fragments that have cross-reacting homologues in female DNA. These cross-reacting sequences, with homologues within both Y and autosomal 2·4 kb fragments, will be referred to as 'common sequences'. In contrast, the larger fragments present in these digests of Y-derived 2·4 kb cloned fragments are only detected with male genomic DNA and are therefore limited to the Y. Thus, like the 3·4 kb *Hae*III fragment (Kunkel & Smith, 1982), the 2·4 kb Y-fragment is composed of both Y-specific and non-Y-specific sequences.

The demonstration that a significant fraction of the 2·4 kb Y repeat is composed of sequences that are specific for the Y is striking, since the sequences within the 2·4 kb Y repeat that have autosomal homologues (common sequences) also detect a 2·4 kb autosomal repeat (Fig. 7). Thus the common sequences are associated with distinguishable Y-specific and autosomal-specific sequences even though the unit size of 2·4 kb is not chromosome-specific. This is unlike the results previously reported for the 3·4 kb *Hae*III repeat, where autosomal homologues yield restriction patterns that are chromosome-specific and distinct from the Y (Burk *et al.* 1985*b*).

Comparisons among several independent Y-2·4 kb *Msp*I clones by cross hybridization (Fig. 9) and restriction mapping (Fig. 10) suggest that there is little variation among the 2·4 kb repeats within the Y chromosome. In these digests, there are no restriction fragments that are unique to any one 2·4 kb Y fragment (Fig. 9) and only a few restriction site differences between cloned fragments (Fig. 10). Comparison of the restriction maps of the cloned fragments with the consensus map derived from analysis of genomic DNA indicates a high degree of similarity among all of the 2·4 kb Y repeats. Although

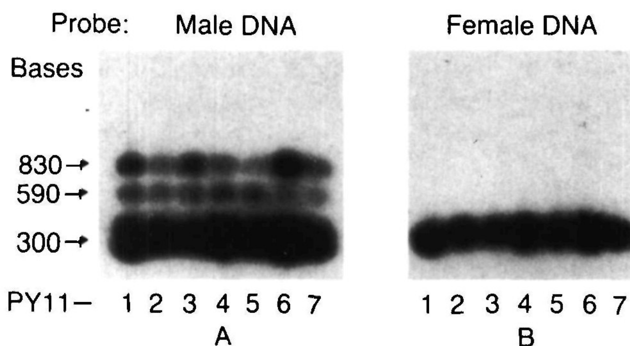

Fig. 8. Hybridization of 2·4 kb Y clones with genomic DNA. Independently derived 2·4 kb *Msp*I repeats from the Y chromosome (pY111–pY117), each of which contains the 2·1 kb *Hae*III fragment, were digested with restriction enzymes *Rsa*I and *Ava*II, and the resulting fragments size fractionated by agarose gel electrophoresis. The digests were hybridized with nick-translated genomic DNA from a male (A) or a female (B) after Southern blotting. Several small bands of approximately 300 bp are hybridized by both male and female genomic DNA probes. The large fragments are only detected with the male probe and are thus confined to the Y chromosome.

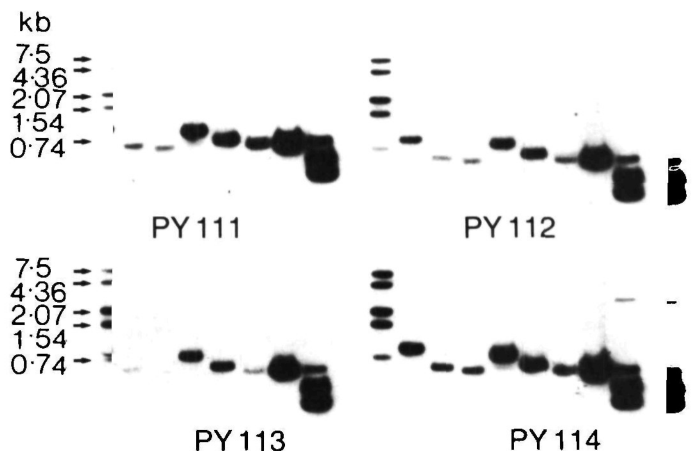

Fig. 9. Cross-hybridization of 2·4 kb Y clones. Recombinant clones of the 2·4 kb Y repeat were each digested with the same series of restriction enzymes and the digests hybridized with radiolabelled Y DNA (pBR322 free) from pY111. Hybridization conditions were as described in Fig. 2 legend. Results are shown here for pY111→pY114, but similar results were obtained with digests of pY115–pY117 and when the other Y clones were used as probes. Radiolabelled DNA size markers are shown on the left of each panel. With the exception of occasional intensity differences, there are no evident differences between these independent Y 2·4 kb repeats.

Fig. 10. Comparison of restriction maps of Y 2·4 kb repeats. Restriction sites were mapped into each of the independently derived Y 2·4 kb clones (pY111–pY117) by standard restriction site mapping methods. The consensus genomic map shown at the bottom was derived by Southern blot analyses of genomic digests probed with 2·4 kb Y repeats. Restriction sites that differ in at least one clone are marked by the arrows; all other sites are present in all fragments. *Acc*I sites were not determined in genomic DNA fragments.

these Y repeats are tandemly arrayed (Fig. 5), they are distributed along the length of the long arm of the Y (Fig. 3) and account for only about 25 % of Y long-arm DNA (Cooke, 1976). Hence, they must occur as independent tandem clusters interspersed among other sequences rather than as a single tandem array. The similarity among all Y 2·4 kb repeats suggests that there is little difference among these repeats in the various tandem clusters scattered along the Y.

We have determined the base sequence of two 2·4 kb *Msp*I fragments from the Y chromosome (data not shown). Comparison of the sequences of these two fragments confirms the extensive homology suggested by restriction analysis (Fig. 10). The two complete sequences determined in our laboratory show little divergence from the partial sequence of a related 2·4 kb element, determined from sequence analysis of *Hin*fI fragments isolated from human satellite I (Frommer, Prosser & Vincent, 1984). The portion of the 2·4 kb Y fragment with autosomal homology (common sequence) is about 800 bp in length, 40 % GC and contains no internal redundancy. As previously reported (Frommer *et al.* 1984), the 2·4 kb Y fragment has a single region that has some homology to the *Alu* repeat family. Our analysis indicates that this *Alu*-like sequence occurs at one end of the common sequence and is flanked by a poly(A) stretch often associated with *Alu* repeats (Jelinek & Schmid, 1982). The *Alu* element in the Y common sequence has about 60 % homology with the *Alu* consensus sequence (Jelinek & Schmid, 1982). The homology is greater with the right half of the *Alu* dimer than with the left half. In this respect, the *Alu*-like sequence within the 2·4 kb Y repeat is similar to the type II *Alu* repeats described in *Galago* (Daniels & Deininger, 1983). Search of the Los Alamos GenBank for homology with the 2·4 kb Y repeat detected no homology except for those with the *Alu* sequences. As predicted from the homology to the consensus *Alu* repeat, extensive homology was seen with the right, but not the left, half of each homologous *Alu* dimer.

It has been proposed that new *Alu* sequences may be generated from transcripts initiated from the RNA polymerase III promoter within the *Alu* element. Thus, it is possible that transcripts initiating in the *Alu*-like sequence within the 2·4 kb repeat might be important in the origin and dispersal of 2·4 kb repeats to various chromosomal sites and help explain the regularity of the 2·4 kb repeat organization. However, direct examination of the Y common *Alu*

sequence failed to detect homology with the consensus *pol*III split promoter sequence (Fowlkes & Shenk, 1980). While the degeneracy of *pol*III promoter sites makes it difficult to exclude a functional *pol*III promoter by sequence analysis, there is at the moment no evidence to suggest that these Y repeats are transcribed by *pol*III.

The Y-specific regions of the 2·4 kb Y repeats are 85% AT. A significant fraction of the Y-specific portion of these repeats is composed of a slightly degenerate 14 bp palindrome. The consensus sequence of this internal repeat is ATAATATA TATTAT.

We have also sequenced an autosomal 2·4 kb fragment. As expected, there is extensive homology between the sequences common to both autosomal and Y repeats. The autosomal common sequence has an *Alu* repeat flanked by a poly(A) stretch in an analogous position to that found on the Y. The autosome-specific sequence is about 80% AT. Although both the Y- and autosome-specific sequences have a high AT content, they do not cross hybridize (Fig. 8) and have only limited sequence homology. It is difficult to devise a means for deriving one from the other by base substitution or sequence rearrangement. Their only common features seem to be their high AT content, their length and their interspersion with the common sequence.

Southern blot hybridization of restriction enzyme digests of female genomic DNA with several cloned autosomal 2·4 kb probes is shown in Fig. 11. Each clone detects the same set of fragments in the various digests, suggesting that the overall genomic organization of each repeat is similar. However, the heterogeneous smear of hybridization seen in most lanes suggests that these autosomal fragments might be more heterogeneous than the analogous fragments from the Y. This was confirmed by direct comparison of individual clones (Fig. 12). In this experiment, five cloned 2·4 kb autosomal fragments were digested with the same restriction enzyme (*Ava*II) and each set of five digests was hybridized with a different autosomal clone after Southern blotting. Each of the autosomal 2·4 kb fragments contains sequences that cross-hybridize with all of the other autosomal 2·4 kb clones. These cross-hybridizing autosomal sequences are the previously discussed common sequences that are homologous to the Y 2·4 kb repeat. In contrast, the higher molecular weight autosome-specific fragments do not cross-hybridize. With the exception of clones pHG243 and 244, which appear to be identical, each clone yields a distinct pattern of autosome-specific fragments which only hybridize with the clone from which they were derived. The low level of hybridization seen with pHG245 is the result of a small amount of common sequence linked to the

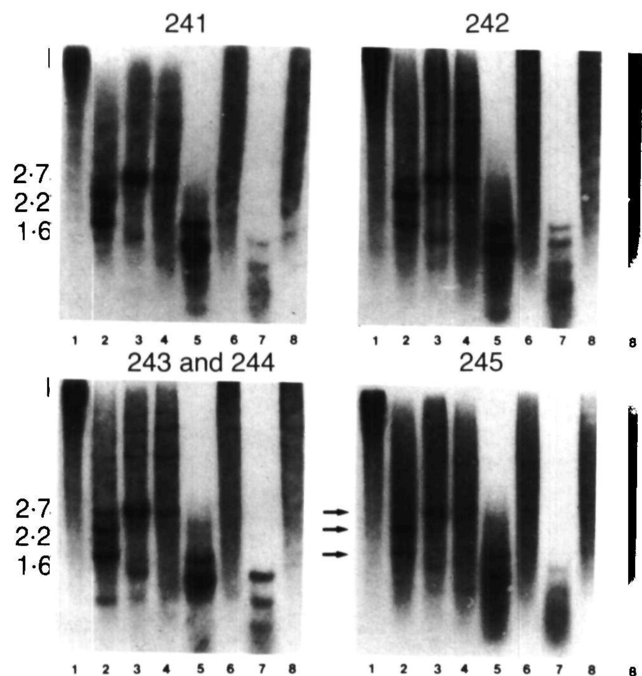

Fig. 11. Restriction digests of female genomic DNA hybridized with autosomal 2·4 kb repeat clones. 5 μg of human female DNA was digested with a series of restriction enzymes and each set of digests was hybridized with one of the radiolabelled 2·4 kb autosomal cloned repeats (241–245) by Southern blotting as described in Fig. 2 legend. No major differences in the prominent genomic bands detected are discernible between clones.

autosome-specific sequence; its self hybridization is much stronger, reflecting its autosome-specific character. Thus, unlike the 2·4 kb Y repeats which are all much the same, the 2·4 kb autosomal repeats contain several distinct autosome-specific sequences, each associated with the same common sequence.

A diagrammatic summary of our analyses of the 2·4 kb tandem repeats is presented in Fig. 13. There are at least four sets of repeats with the same overall organization. On the Y chromosome, common sequences of about 0·8 kb are interspersed with 1·6 kb sequences which are confined to the Y. The majority of the repeats on the Y are composed of the same common and Y-specific sequences. All of the autosomal repeats have the same common sequences which differ from those on the Y by the presence of an additional *Hae*III site. This *Hae*III site accounts for the smaller unit size of the autosomal *Hae*III repeats (i.e. 1·9 kb for the autosomal *Hae*III repeats compared to 2·1 kb for the Y *Hae*III repeats). Presumably, the common sequence accounts for the similarity in genomic restriction patterns seen in Fig. 11. The autosomal 2·4 kb repeats have at least four different autosome-specific sequences. These autosome-specific sequences (1·6 kb) alternate with the autosomal common sequences (0·8 kb) to yield

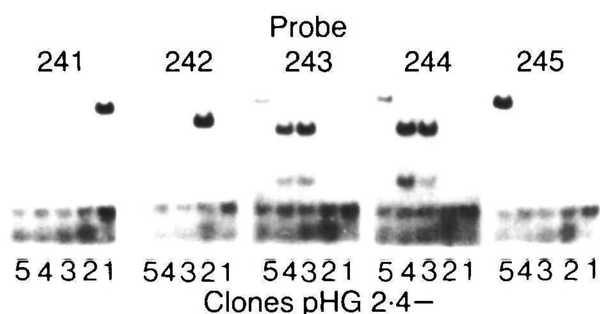

Fig. 12. Cross-hybridization of 2·4 kb autosomal clones. Autosomal 2·4 kb *Pvu*II fragments homologous to the 2·4 kb Y repeat were cloned into pBR322. Each clone in the series (pHG241 → 245) was digested with the restriction enzyme *Ava*II. Each set of digests was hybridized with one of the radiolabelled purified human DNA inserts from the 2·4 kb autosomal clones. With the exception of pHG243 and pHG244, each clone has at least one restriction fragment which is unique. Hybridization with a 2·4 kb Y repeat detects the pair of low molecular weight fragments which also cross-react between the autosomal clones. Thus, these autosomal clones contain common sequences which have homology in both autosomal and Y repeats. Each autosomal clone, except pHG243 and 244, which appear to be the same, has an autosome-specific sequence which does not have homology with the Y repeats or the other autosomal clones shown here.

structures analogous to the Y-specific fragments. Thus, the common sequences appear capable of associating with a variety of sequences in such a way that identical tandem repeat units are formed. These findings are consistent with the notion that these common sequences have a degree of mobility within the genome and that the mechanism by which they move determines a particular structural organization.

This notion of mobility is supported by studies of gorilla sequences homologous to the human 2·4 kb repeats. Fig. 14 compares various restriction digests of human and gorilla DNA probed with a 2·4 kb Y fragment. In contrast to results with human DNA (lower panel), the pattern and hybridization intensity of restriction fragments seen in male and female gorilla DNA (upper panel) are identical. This suggests that gorilla 2·4 kb repeat homologues are primarily autosomal and are not present on the gorilla Y chromosome. This has been confirmed by re-association kinetics and *in situ* hybridization. Since homologues of the Y-specific segment of the 2·4 kb Y fragment are not detected in gorilla DNA (data not shown), the hybridizing gorilla fragments must contain homologues of the common sequence. In addition to the demonstrated cross-hybridization, two indirect lines of evidence support the notion that the 2·4 kb common homologues are similar in the two

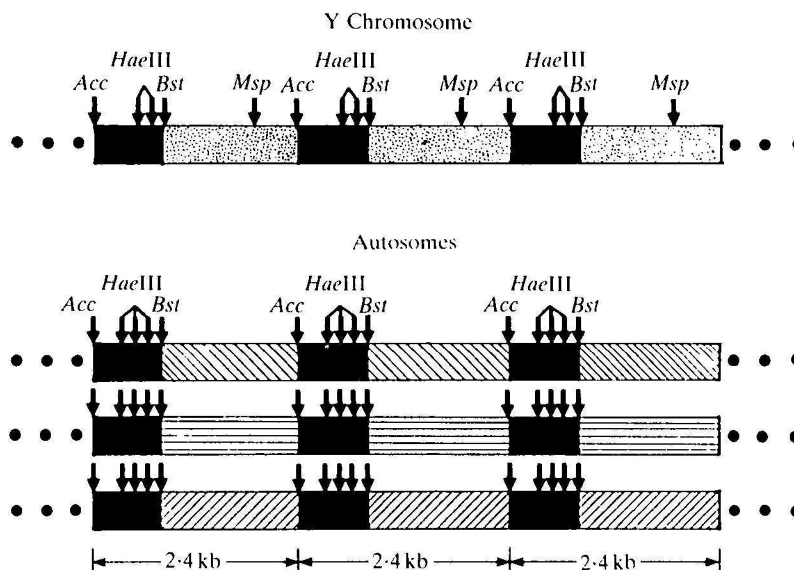

Fig. 13. Model of Y and autosomal 2·4 kb repeat organization. Each 2·4 kb repeat is composed of tandem repeats of common and chromosome-specific sequences interspersed in a regularly repeating manner such that the overall length of the repeating unit is 2·4 kb. At all chromosomal sites, 1·6 kb chromosome-specific sequences (striped or stippled) are flanked by 800 bp common sequences (black). The common sequences are direct repeats with *Acc*I and *Bst*NI restriction sites at or near their boundaries. The common and Y-specific sequences on the Y show little sequence divergence among themselves and each tandem cluster contains similar interspersed common and Y-specific sequences. All of the autosomal repeats have the same common sequence which differs from the Y common sequences by a few restriction sites (such as an additional *Hae*III site in the autosomal common sequence). The autosomal common sequence is also a direct repeat bordered by *Acc*I and *Bst*NI restriction sites. There are at least four different autosome-specific sequences associated with the autosomal tandem clusters.

species. First, the 2·4 kb autosomal common sequence in humans has a 700 bp fragment flanked by *Hinf*I and *Bst*NI restriction sites. Digestion of gorilla DNA with these two enzymes yields a fragment of about 700 bp with homology to the 2·4 kb common sequence (Fig. 15). In addition, a 790 bp fragment can be seen in this figure which also hybridizes with the 2·4 kb common sequence. An equivalent fragment is not seen in similar digests of human DNA. This suggests that, as in humans, gorillas may have more than one common sequence. Second, restriction enzymes with sites in the 2·4 kb common sequence that yield 2·4 kb fragments from human DNA also yield distinct fragments, of 2·7 and 3·0 kb, in gorilla (Fig. 16). Direct comparison of the similarity of

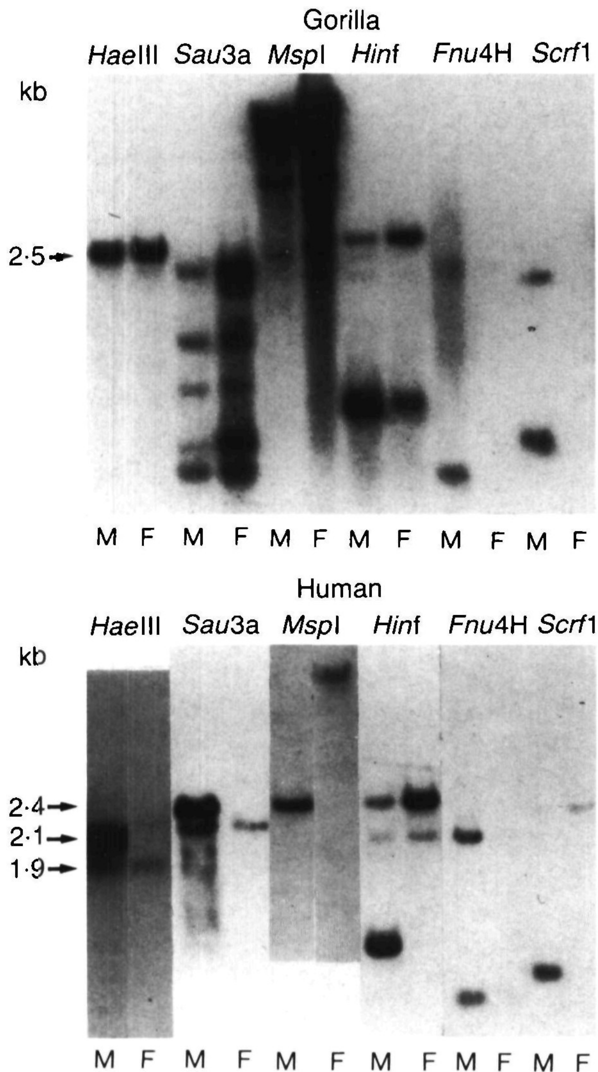

Fig. 14. Primate restriction fragments with 2·4 kb Y repeat homology. Genomic DNA from male (M) and female (F) gorillas and humans were digested with the restriction enzymes indicated at the top of each panel. Each set of digests was hybridized with the same 2·4 kb Y repeat by the methods indicated in Fig. 2. The hybridized bands reflect 2·4 kb common sequence homology since the Y-specific 2·4 kb sequences do not hybridize with gorilla DNA under these hybridization conditions. The differences in intensity of the hybridized bands between human males and females reflects the concentration of homologues on the Y chromosome. The similarity in the extent of reaction with male and female gorilla DNA indicates that the gorilla homologues are not present on the gorilla Y.

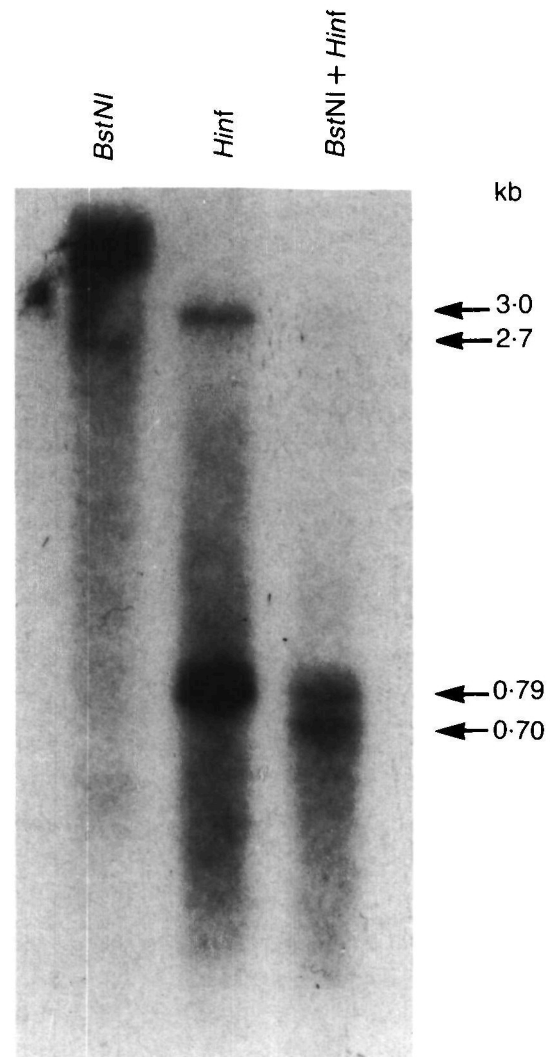

Fig. 15. Length estimates of gorilla 2·4 kb Y common sequence homologues. Sequence analysis and restriction site mapping predict that a 700 bp 2·4 kb common sequence is generated from human autosomal 2·4 kb repeats by digestion with *Hinf*I and *Bst*NI. Digestion of gorilla DNA with these same enzymes also reveals a 700 bp fragment with 2·4 kb common sequence homology. A 790 bp fragment is also seen in the double digest as well as in the *Hinf*I control digest. This fragment is not seen in similar digests of either male or female human DNA. DNA fragments in these digests were separated on a 1·4 % agarose gel using sodium borate electrophoresis buffer to maximize separation of fragments in the 500–1000 bp range. Hybridization conditions were those described in the legend of Fig. 2.

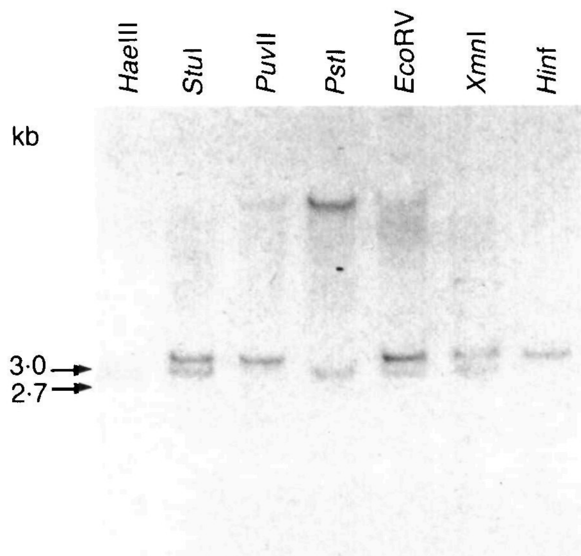

Fig. 16. Gorilla restriction fragments homologous with the human 2·4 kb Y common sequence. Gorilla DNA (5 μg/lane) was digested with the restriction enzymes indicated at the top of the figure. The digests were hybridized after Southern blotting with a radiolabelled Y common sequence subcloned from a 2·4 kb Y repeat.

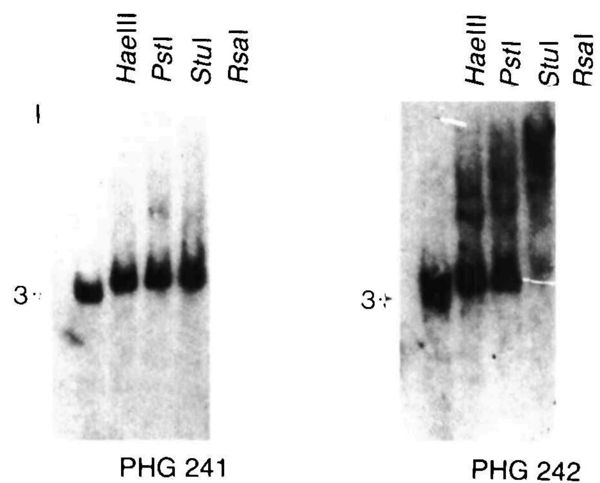

Fig. 17. Gorilla restriction fragments homologous with human 2·4 kb autosome-specific sequences. Gorilla DNA (5 μg/lane) was digested with the restriction enzymes indicated at the top of the figure. After transfer to nitrocellulose filters, the digests were hybridized with radiolabelled autosome-specific subclones prepared from two different 2·4 kb autosomal repeats, pHG241 and pHG242. The homologous fragments detected are distinct from those observed with the common sequence probe derived from the 2·4 kb Y repeat (Fig. 16). In human, these autosome-specific subclones hybridize to 1·9 kb fragments in *Hae*III digests and to 2·4 kb fragments in the other digests. As seen, the pattern of gorilla restriction fragments detected by these two probes is different.

human and gorilla common sequences requires analysis of cloned sequences. However, the gorilla repeat has proved to be difficult to clone and such a comparison cannot yet be made.

The presence of two fragments (2·7 and 3·0 kb), containing a 700 or 790 bp sequence homologous to the common sequence of the human 2·4 kb repeat, suggests that, as in humans, the gorilla common sequence may be associated with more than one non-common sequence. We examined gorilla DNA to determine if homologues of the autosome-specific sequences of human 2·4 kb repeats are associated with the gorilla common sequences. Hybridization of gorilla DNA with subclones of the human 2·4 kb fragment containing only the autosomal-specific sequences demonstrates that these sequences also have identifiable repeat structures in the gorilla (Fig. 17). The repeat length of these sequences is different from that detected in human DNA and from the repeats in gorilla detected with the common sequence (Fig. 16). Thus, while both components of the human autosomal 2·4 kb repeats (the common and autosome-specific sequences) are present in the gorilla, they are not associated with each other. These observations provide additional evidence that the common sequences are capable of associating with a variety of sequences in such a way that precisely defined repeating units are generated.

The structure of the genomic repeating elements homologous with the Y 3·4 kb *Hae*III fragments also suggests that it may contain sequences allowing for genomic mobility. Like the 2·4 kb repeat, the 3·4 kb *Hae*III fragments contain Y-specific sequences interspersed with non-Y-specific ones. However, unlike the 2·4 kb Y repeats, the 3·4 kb Y fragments are heterogeneous with several non-Y-specific and many Y-specific sequences (Kunkel *et al.* 1979). In addition, there are several distinguishable autosomal domains which contain homologues of the 3·4 kb repeat. While the Y-chromosome fragments contain all of the various sequences common to the Y and autosomes, each autosomal domain with 3·4 kb common sequence homology contains just one common sequence (Burk *et al.* 1985*b*). Thus the organizational heterogeneity of the various autosomal sequences with Y 3·4 kb homology is much greater than those with 2·4 kb Y-fragment homology. Based on our published work (Burk *et al.* 1985*b*) and analyses of cloned Y 3·4 kb fragments, a model of the organization of Y 3·4 kb fragments can be constructed. As shown in Fig. 18, tandem clusters of these repeats, containing a heterogeneous mixture of sequences, occur on the Y. Their distribution throughout the length of the Y long arm indicates the existence of several separate tandem

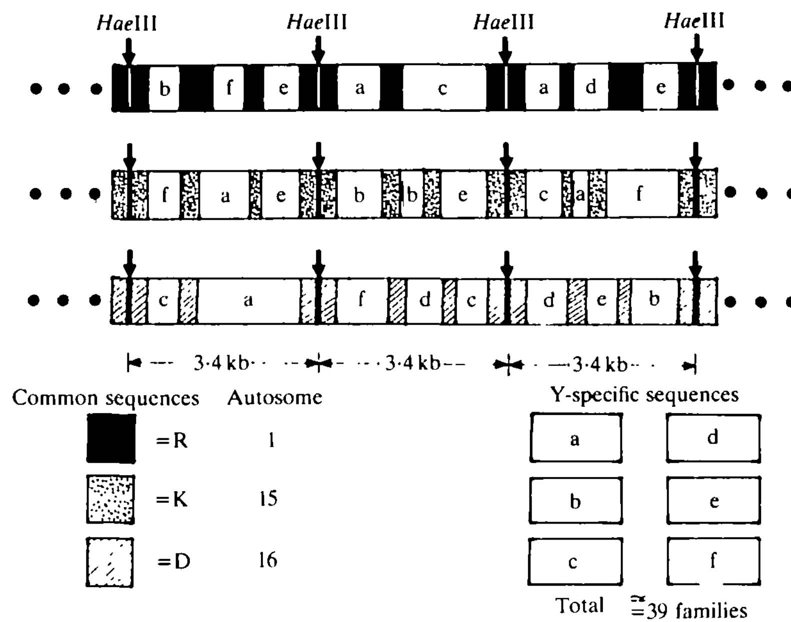

Fig. 18. Diagrammatic representation of 3·4 kb Y tandem repeat organization. The three heterogeneous clusters of repeats shown represent some of the organizations found on the Y chromosome. They are based on the analysis of genomic DNA and cloned examples of this repeat. There are at least three non-homologous common sequences among the Y repeats (black, stippled or striped). Different clusters of Y repeats have different common sequences, each of which is homologous to a chromosome-specific autosomal domain (shown in the lower left of the figure). Each Y cluster is characterized by a common sequence containing a *Hae*III site at regular 3·4 kb intervals. Additional copies of the common sequences occur within the 3·4 kb repeating unit with variable number, location and length. The common sequences are interspersed on the Y chromosome with a number of different Y-specific repeats (white; also shown in the lower right of the picture).

clusters. All tandem clusters on the Y are characterized by a common set of *Hae*III sites at 3·4 kb intervals. Based on analyses of individual cloned 3·4 kb *Hae*III fragments, each 3·4 kb fragment contains a particular non-Y-specific sequence interspersed with a number of different Y-specific sequences. The position and length of the internal non-Y-specific sequences are variable with respect to length and position. Each non-Y-specific sequence has homology with a distinct autosomal domain. The structural characteristics of these fragments are reminiscent of the scrambled repeats and some transposable elements in *Drosophila* (Rubin, 1983).

Conclusion

Most of DNA sequences within the human Y-chromosome are not associated with the Y chromosome of other primates. Recent transpositions of single copy DNA from the X chromosome (Page *et al.* 1984; Cooke *et al.* 1984) and autosomes (Burk *et al.* 1985*a*) to the human Y chromosome have been reported. In at least one case, it has been documented that DNA sequences from the non-pairing region of the human Y chromosome that have X-chromosome homology in humans have homology with the X but not the Y chromosome in other higher primates (Page *et al.* 1984). In addition, as presented in this report, repeated DNA sequences that are specific for the human Y chromosome (Kunkel & Smith, 1982) or have specific organizational characteristics on the human Y are not present on the Y chromosome of other primates (Cooke, Schmidtke & Gosden, 1982). Rather, those human Y-chromosome repeat sequences that have homology with non-human primates detect homologous autosomal sequences. Although the chromosomal location of these Y-repeat homologues is not known in non-human primates, they map to a few specific autosomal sites in humans. The major autosomal regions with 3·4 kb *Hae*III Y-repeat homology have been mapped to chromosomes 1, 9, 15, 16, 21 and 22 (Burk *et al.* 1985*b*), while the autosomal 2·4 kb repeats have a major concentration on chromosome 14. Since these major Y-chromosome tandem repeats are not known to be transcribed, it is unlikely that they have become associated with the human Y chromosome by a mechanism involving an RNA intermediate. It seems more likely that a significant fraction of the human Y chromosome originated by a series of translocations or transpositions involving the X chromosome and several autosomes. In this regard, it should be noted

(Burk *et al.* 1985*b*) that the chromosomal distribution of the 3·4 kb *Hae*III Y-repeat autosomal homologues corresponds to autosomal sites that are frequently involved in Y-chromosome long-arm translocations (Smith, Fraser & Elliot, 1979) and are frequently associated with both homologous and heterologous somatic pairing (Ford, Callen, Roberts & Jahnke, 1983; Schmid, Grunert, Haaf & Engel, 1983).

If multiple transposition events were involved in the generation of the human Y chromosome, they must have been followed by selective amplification of particular sequences which resulted in the variety of Y-chromosome tandem repeats described here. The heterogeneity of sequences within each family of tandem repeats suggests that such amplification events must have occurred several times. The occurrence of similar clusters of each family of tandem repeats dispersed along the Y chromosome, with each cluster having the same overall organization, suggests that the events giving rise to the various Y repeats must have occurred at several independent sites. Alternatively, these repeats may have been generated as single tandem arrays and subsequently shuffled to various sites along the Y long arm, perhaps as a result of an inherent instability of the non-pairing region of the Y. The possibility of the linear instability of the Y chromosome has previously been suggested (Affara *et al.* 1986). Whatever the mechanism, the regularity in the organization of these repeats, which contain heterogeneous collections of sequences, and their specific chromosomal organizations in both humans and related primates, make it unlikely that they arose from purely stochastic processes (Dover, 1982) and suggest that they may play some role in the function of segregation of the Y chromosome. This notion would be strengthened if it could be demonstrated that sequences with structural characteristics similar to those described here for the human Y chromosome also occur in the Y chromosome of non-human primates.

We would like to thank Drs L. M. Kunkel, H. F. Willard and S. H. Boyer for their valuable contributions to much of the work presented here. We also thank Ms C. J. Decker and Ms C. T. Comey for their helpful discussion of the manuscript and Ms C. Dove for help in preparing the manuscript. This work was supported in part by NIH grant HD 17161 (KDS) and the Howard Hughes Medical Institute.

References

ADAMS, J. W., KAUFMAN, R. E., KRETSCHMER, P. S., HARRISON, M. & NIENHUIS, Q. W. (1980). A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. *Nucl. Acids Res.* **8**, 6113–6128.

AFFARA, N. A., FERGUSON-SMITH, M. A., TOLMIE, J., KWOK, K., MITCHELL, M., JAMIESON, D., COOKE, A. & FLORENTIN, L. (1986). Variable transfer of Y-specific sequences in XX males. *Nucl. Acids Res.* **14**, 5375–5387.

BRITTEN, R. J. & DAVIDSON, E. H. (1969). Gene regulation for higher cells: a theory. *Science* **165**, 349–357.

BRITTEN, R. J. & DAVIDSON, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46**, 111–138.

BROWN, S. D. M. & DOVER, G. A. (1980). The specific organization of satellite DNA sequences on the X chromosome of musculus: partial independence of chromosome evolution. *Nucl. Acids Res.* **8**, 781–792.

BRUTLAG, D. L. (1980). Molecular arrangement and evolution of heterochromatic DNA. *A. Rev. Genet.* **14**, 121–144.

BURK, R. D., PATRICK, M. A. & SMITH, K. D. (1985*a*). Characterization and evolution of a single-copy sequence from the human Y chromosome. *Molec. cell Biol.* **5**, 576–581.

BURK, R. D., SZABO, P., O'BRIEN, S., NASH, W. G., YU, L. & SMITH, K. D. (1985*b*). Organization and chromosomal specificity of autosomal homologs of human Y chromosome repeated DNA. *Chromosoma* **92**, 225–233.

COOKE, H. (1976). Repeated sequence specific to human males. *Nature, Lond.* **262**, 182–186.

COOKE, H. J., BROWN, W. A. R. & RAPPOLD, G. A. (1984). Closely related sequences on human X and Y chromosomes outside the pairing region. *Nature, Lond.* **311**, 259–261.

COOKE, H. J., SCHMIDTKE, J. & GOSDEN, J. R. (1982). Characterization of a human Y chromosome repeated sequence and related sequences in higher primates. *Chromosoma* **87**, 491–502.

DANIELS, G. R. & DEININGER, P. L. (1983). A second major class of *Alu* family repeated DNA sequences in a primate genome. *Nucl. Acids Res.* **11**, 7595–7610.

DAVIDSON, E. H. & BRITTEN, R. J. (1979). Regulation of gene expression: possible role of repetitive sequences. *Science* **204**, 1052–1059.

DEININGER, P. H., JOLLY, D. J., RUBIN, C. M., FRIEDMANN, T. & SCHMID, C. W. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. *J. molec. Biol.* **151**, 17–33.

DOOLITTLE, W. F. & SAPIENZA, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature, Lond.* **284**, 601–603.

DORMAN, B. P., SHIMIZU, N. & RUDDLE, F. H. (1978). Genetic analysis of the human cell surface: Antigenic marker for the human X chromosome in human-mouse hybrids. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2363–2367.

DOVER, G. (1982). Molecular drive: A cohesive mode of species evolution. *Nature, Lond.* **299**, 111–117.

DOVER, G. A. & FLAVELL, R. B. (1984). Molecular coevolution: DNA divergence and the maintenance of function. *Cell* **38**, 622–623.

DUNCAN, C., BIRO, P. A., CHOUDARY, P. V., ELDER, J. T., WANG, R. R. C., FORGET, B. G., DERIEL, J. K. & WEISSMAN, S. M. (1979). RNA polymerase III transcriptional units are interspersed among human non-α-globin genes. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5095–5099.

FORD, J. H., CALLEN, D. F., ROBERTS, C. G. & JAHNKE, A. B. (1983). Interactions between C-bands of chromosomes 1 and 9 in recurrent reproductive loss. *Hum. Genet.* **63**, 58–62.

FOWLKES, D. M. & SHENK, T. (1980). Transcriptional control regions of the adenovirus VAI RNA gene. *Cell* **22**, 405–413.

FRITSCH, E. F., LAWN, R. M. & MANIATIS, T. (1980). Molecular cloning and characterization of the human beta-like globin gene cluster. *Cell* **19**, 959–972.

FROMMER, M., PROSSER, J. & VINCENT, P. C. (1984). Human satellite I sequences include a male specific 2·47 kb tandemly repeated unit containing one *Alu* family member per repeat. *Nucl. Acids Res.* **12**, 2887–2900.

GALAU, G. A., KLEIN, W. H., DAVIS, M. M., WOLD, B. J., BRITTEN, R. J. & DAVIDSON, E. H. (1976). Structural gene sets active in embryos and adult tissues of the sea urchin. *Cell* **7**, 487–505.

GOODFELLOW, P., DARLING, S. & WOLFE, J. (1985). The human Y chromosome. *J. med. Genet.* **22**, 329–344.

GRIMALDI, G. & SINGER, M. F. (1983). Members of the *Kpn* I family of long interspersed repeated sequences join and interrupt α-satellite in the monkey genome. *Nucl. Acids Res.* **11**, 321–338.

GRIMALDI, G., SKOWRONSKI, J. & SINGER, M. F. (1984). Defining the beginning and end of *Kpn* I family segments. *EMBO J.* **3**, 1753–1759.

HWU, H. R., ROBERTS, J. W., DAVIDSON, E. H. & BRITTEN, R. J. (1986). Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3875–3879

JAGADEESWARAN, P., FORGET, B. G. & WEISSMAN, S. M. (1981). Short interspersed repetitive DNA elements in eucaryotes: Transposable DNA elements generated by reverse transcription of RNA pol III. *Cell* **26**, 141–142.

JELINEK, W. R., TOOMEY, T. P., LEINWAND, L., DUNCAN, C. H., BIRO, P. A., CHOUDARY, P. V., WEISSMAN, S. M., RUBIN, C. M., HOUCK, C. M., DEININGER, P. L. & SCHMID, C. W. (1980). Ubiquitous interspersed repeated sequences in mammalian genomes. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1398–1402.

JELINEK, W. R. & SCHMID, C. W. (1982). Repetitive sequences in eukaryotic DNA and their expression. *A. Rev. Biochem.* **51**, 813–844.

KOLE, L. B., HAYNES, S. R. & JELINEK, W. R. (1983). Discrete and heterogeneous high molecular weight RNAs complementary to a long dispersed repeat family (a possible transposon) of human DNA. *J. molec. Biol.* **165**, 257–286.

KUNKEL, L. M., SMITH, K. D. & BOYER, S. H. (1979). The organization and heterogeneity of sequences within a repeating unit of human Y chromosome DNA. *Biochemistry* **18**, 3343–3352.

KUNKEL, L. M. & SMITH, K. D. (1982). Evolution of human Y-chromosome DNA. *Chromosoma* **86**, 209–228.

LEE, T. N. & SINGER, M. F. (1982). Structural organization of alpha-satellite DNA in a single monkey chromosome. *J. molec. Biol.* **161**, 323–342.

LEE, T. N. H. & SINGER, M. F. (1986). Analysis of LINE-1 family sequences on a single monkey chromosome. *Nucl. Acids Res.* **14**, 3859–3870.

MARTIN, S. L., VOLIVA, C. F., HARDIES, S. C., EDGELL, M. H. & HUTCHINSON III, C. A. (1985). Tempo and mode of concerted evolution in the L1 repeat family of mice. *Mol. Biol. Evol.* **2**, 127–140.

MANUELIDIS, L. (1982). Repeated DNA sequences and nuclear structure. In *Genome Evolution* (ed. G. A. Dover & R. B. Flavell), pp. 263–285. New York: Academic Press.

MANUELIDIS, L. & BIRO, P. A. (1982). Genomic representation of the *Hin*dIII 1·9 kb repeated DNA. *Nucl. Acids Res.* **10**, 3221–3239.

MANUELIDIS, L. & WARD, D. C. (1984). Chromosomal and nuclear distribution of the *Hin*dIII 1·9 kb human DNA repeat segment. *Chromosoma* **91**, 28–38.

MIKLOS, G. L. G. & JOHN, B. (1979). Heterochromatin and satellite DNA in man: properties and prospects. *Am. J. Hum. Genet.* **31**, 264–280.

MIYAKE, T., MIGITA, K. & SAKAKI, Y. (1983). Some *Kpn* I family members are associated with the *Alu* family in the human genome. *Nucl. Acids Res.* **11**, 6837–6845.

O'BRIEN, S. J., BONNER, T. I., COHEN, M., O'CONNELL, C. & NASH, W. G. (1983). Mapping of an endogenous retroviral sequence to human chromosome 18. *Nature, Lond.* **303**, 74–77.

ORGEL, L. E. & CRICK, F. H. (1980). Selfish DNA: the ultimate parasite. *Nature, Lond.* **284**, 604–607.

PAGE, D. C., HARPER, M. E., LOVE, J. & BOTSTEIN, D. (1984). Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature, Lond.* **311**, 119–122.

POTTER, S. S. (1984). Rearranged sequences of a human *Kpn* I element. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1016–1021.

RINEHART, F. P., RITCH, T. G., DEININGER, P. L. & SCHMID, C. W. (1981). Renaturation rate studies of a single family of interspersed repeated sequences in human deoxyribonucleic acid. *Biochemistry* **20**, 3003–3010.

RUBIN, G M. (1983). Dispersed Repetitive DNAs in *Drosophila*. In *Mobile Genetic Elements* (ed. J. A. Shapiro), pp. 329–361. New York: Academic Press.

SCHMECKPEPER, B. J., WILLARD, H. F. & SMITH, K. D. (1981) Isolation and characterization of cloned human DNA fragments carrying reiterated sequences common to both autosomes and the X chromosome. *Nucl. Acids Res.* **8**, 1853–1872.

SCHMECKPEPER, B. J., SCOTT, A. F. & SMITH, K. D. (1984). Transcripts homologous to a long repeated DNA element in the human genome. *J. biol. Chem.* **259**, 1218–1225.

SCHMID, C. W., GRUNERT, D., HAAF, T. & ENGEL, W. (1983). A direct demonstration of somatically paired

heterochromatin of human chromosomes. *Cytogenet Cell Genet.* **36**, 554–561.

SCHMID, C W. & JELINEK, W R. (1982) The Alu family of dispersed repetitive sequences. *Science* **216**, 1065–1070.

SHAFIT-ZAGARDO, B , MAIO, J J. & BROWN, F. L. (1982). *Kpn* I families of long, interspersed repetitive DNAs in human and other primates genomes. *Nucl. Acids Res.* **10**, 3175–3193.

SHAFIT-ZAGARDO, B., BROWN, F L., ZAVODNY, P J. & MAIO, J. J (1983). Transcription of the *Kpn* I families of long interspersed DNAs in human cells. *Nature, Lond.* **304**, 277–280.

SIMMLER, M. C., ROUYER, F , VERGNAUD, G., NYSTROM-LAHTI, M., YEN NGO, K., DE LA CHAPELLE, A. & WEISSENBACH, J. (1985). Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature, Lond.* **317**, 692–697.

SINGER, M. F. (1982). Highly repeated sequences in mammalian genomes. *Int. Rev. Cytol* **76**, 67–112.

SMITH, A., FRASER, I S. & ELLIOT, G. (1979). An infertile male with balanced Y-19 translocation. Review of Y-autosome translocations. *Ann. Génét.* **22**, 189–194.

SZABO, P., KUNKEL, L. M., YU, L. C., GEORGE, D. & SMITH, K. D. (1980). Chromosomal distribution of DNA sequences derived from the human Y chromosome in human and higher primates *Cytogenet Cell Genet.* **25**, 212–213.

ULLU, E & TSCHUDI, C. (1984). *Alu* sequences are processed 7SL RNA genes *Nature, Lond.* **312**, 171–172.

ULLU, E. & WEINER, A. M (1985) Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature, Lond.* **318**, 371–374.

VAN ARSDELL, S. W., DENISON, R. A., BERSTEIN, L. B., WEINER, A. M., MANSER, T. & GESTELAND, R. F. (1981). Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell* **26**, 11–17.

VOLIVA, C. F., JAHN, C. L., COMER, M. B., HUTCHISON III, C. A. & EDGELL, M. H. (1983). The LIMd long interspersed repeat family in the mouse: almost all examples are truncated at one end. *Nucl Acids Res.* **11**, 8847–8859

WILLARD, H. F., SMITH, K. D. & SUTHERLAND, J. (1983). Isolation and characterization of a major tandem repeat family from the human X chromosome. *Nucl Acids Res.* **11**, 2017–2033.

WOLFE, J , ERICKSON, R. P., RIGBY, P. W. J. & GOODFELLOW, P. N. (1984). Cosmid clones derived from both euchromatic and heterochromatic regions of the human Y chromosome. *EMBO J.* **3**, 1997–2003.

WOLFE, J., DARLING, S. M., ERICKSON, R. P., CRAIG, I. W., BUCKLE, V. J., RIGBY, P. W. J., WILLARD, H. F. & GOODFELLOW, P. N. (1985). Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. *J. molec. Biol* **82**, 477–485.

YOUNG, K. E., WILLARD, H. F. & SMITH, K. D (1983) Molecular and chromosomal organization and analysis of DNA methylation of the human Y chromosome specific 2·1 kb *Hae*III DNA fragment. *Am. J. Hum. Genet.* **33**, 60A.