The Human Y Chromosome: Overlapping DNA Clones Spanning the Euchromatic Region

Simon Foote, Douglas Vollrath, Adrienne Hilton, David C. Page

The human Y chromosome was physically mapped by assembling 196 recombinant DNA clones, each containing a segment of the chromosome, into a single overlapping array. This array included more than 98 percent of the euchromatic portion of the Y chromosome. First, a library of yeast artificial chromosome (YAC) clones was prepared from the genomic DNA of a human XYYYY male. The library was screened to identify clones containing 160 sequence-tagged sites and the map was then constructed from this information. In all, 207 Y-chromosomal DNA loci were assigned to 127 ordered intervals on the basis of their presence or absence in the YAC's, yielding ordered landmarks at an average spacing of 220 kilobases across the euchromatic region. The map reveals that Y-chromosomal genes are scattered among a patchwork of X-homologous, Y-specific repetitive, and single-copy DNA sequences. This map of overlapping clones and ordered, densely spaced markers should accelerate studies of the chromosome.

Complete physical maps, consisting of overlapping recombinant DNA clones spanning an entire genome, are a primary guide for exploring the arrangement of an organism's genetic material and the information it encodes. Such physical maps facilitate correlation of genetic linkage maps and chromosome banding patterns with the underlying DNA, offer immediate access to clones from any region, and provide substrates for large-scale nucleotide sequencing. Traditional DNA cloning vectors such as phage and cosmids have been used to construct complete or nearly complete physical maps of genomes for several organisms, including the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae*, and the nematode *Caenorhabdites elegans* (1). With the advent of yeast artificial chromosome (YAC) vectors capable of propagating much larger pieces of DNA (2), it has become feasible to consider assembling a similar map of the human genome. Indeed, YAC-based physical maps have been constructed for the genomes of several lower organisms (3), as well as for the q24-q28 region of the human X chromosome (4).

We now report the construction of an essentially complete physical map of a human chromosome, the Y chromosome. The Y chromosome is an appropriate target for physical mapping at this time. First, genetic mapping is impossible except in the small pseudoautosomal region, the only part of the Y chromosome that undergoes meiotic recombination. Second, the Y is one of the smallest of human chromosomes, with an estimated average size of 60 million base pairs (Mb) (5). Cytologically, the human Y chromosome consists of a heterochromatic region and a euchromatic region. The heterochromatic region is situated on the distal long arm (Yq) and varies in size, in that it constitutes more than half of the chromosome in some normal males but is virtually undetectable in others (6). Composed of highly repeated sequences (*DYZ1* and *DYZ2*), the heterochromatic region has an estimated sequence complexity of less than 10 kb (7). The euchromatic region—the short arm (Yp), centromere, and proximal long arm—constitutes the remainder of the chromosome, and its size is quite constant among normal males. The euchromatic region contains blocks of sequences homologous to the X chromosome, families of Y-specific repetitive sequences, and all genes identified on the Y chromosome. In this article we describe the physical mapping of the euchromatic region.

The physical map was assembled by a procedure called STS (sequence-tagged sites) content mapping (8). In our case, we screened a human genomic YAC library using polymerase chain reaction (PCR) assays (9) to identify the clones containing 160 STS's (10) from the Y chromosome. Overlap between YAC clones was evidenced by their having such sites in common. The approach has the advantage of simultaneously isolating specific YAC's from a complex library, ordering them into contigs, and arranging the sites into a finely ordered set of points along the chromosome. Vollrath *et al.* (11) have described the placement of the STS's used in our study into 43 ordered intervals by deletion mapping. This prior information of the approximate order of STS's simplified the analysis in assembling YAC contigs (overlapping arrays). Moreover, because the STS's were all known to lie on the Y chromosome and most were roughly ordered, we avoided problems posed by "chimeric" YAC's, that is, artifactual clones that contain DNA from two or more different genomic regions and that constitute a sizable fraction of many libraries (12).

Y-chromosomal YAC's. Assembling a set of YAC's sufficient to span a human chromosome requires a YAC library with substantial redundancy of coverage (13), preferably with large inserts. Given these considerations, we constructed a YAC library (14) using a human cell line (OXEN) derived from an XYYYY male (15), and thereby obtained a representation of the Y chromosome four times larger than would have resulted had the DNA from a normal XY male been used. The library consisted of 10,368 clones with an average insert size of 650 kilobases (kb), and we estimated that each point of the Y chromosome was sampled about 4.5 times.

For practical screening of the YAC library for 160 STS's, identification of individual YAC's by PCR must be efficient. We therefore chose a hierarchical three-step screening system. In the first step, each STS was analyzed on 18 pools, each containing 576 YAC's. The second and third steps were assays on subpools which were performed only when the previous step had yielded a positive result (16). On average, about 25 PCR assays were sufficient to assign an STS to a particular YAC.

In all, 234 Y chromosomal YAC's were isolated from the library, and these had a mean size of 580 kb (SD ± 253 kb). The sizes were determined by pulsed-field gel electrophoresis (PFGE) (17) designed to separate the yeast chromosomes, Southern blot transfer to fix the DNA, and hybridization of the blot (18) with human DNA to visualize the YAC. In some cases, DNA from a single yeast colony showed multiple bands hybridizing to human DNA, suggesting that the YAC contained human sequences that were undergoing deletion during propagation in yeast. In addition, YAC's from certain regions of the Y chromosome were noticeably smaller than the average, suggesting that those had already sustained deletions. Particularly unstable were regions containing known tandem repeats such as *DYZ3* (alphoid) and *DYZ4* sequences (19), the pseudoautosomal region, and a portion of Yp homologous to Xq21.

Of the 160 STS's, four were not represented in the library. Two of these four, sY1 and sY2, are immediately subtelomeric (20). Their absence was not surprising since

The authors are at the Howard Hughes Research Laboratories at Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142.

the library was constructed by methods biased against subtelomeric regions. A third STS, sY162, is X-Y homologous and may also be subtelomeric; the fourth, sY7, is pseudoautosomal but not subtelomeric. As is discussed later, we suspect that sY7 may tend to undergo deletion from YAC's.

Assembling YAC contigs by STS content. When the library screening was completed, the STS content of each Y chromosomal YAC was known. This information was used to construct contigs, on the basis of the premise that overlapping YAC's may share an STS, whereas nonoverlapping YAC's will not (Fig. 1). Progressive incorporation of more and more of the STS content data resulted in small arrays of overlapping YAC's being fused into larger contigs. This procedure alone made it possible to assemble 196 Y chromosomal YAC's into ten contigs separated by nine gaps.

This assembly of contigs was simplified because all of the STS's were known to map to the Y-chromosome and most had been roughly ordered by deletion mapping (11). This prior ordering made any sophisticated computation unnecessary. However, our experience suggests that contigs can be readily assembled from a collection of STS's where only some have been ordered by other means, as in the case of a large block of Yp that is homologous to Xq21 (Fig. 2) (11). Although

17 of the 33 STS's from this Xq21-homologous region had not been ordered by deletion mapping, all 33 STS's were useful in assembling YAC contigs, and all 33 STS's were unambiguously placed on the map.

The chief complication in contig assembly was that some of the STS's corresponded not to single-copy sequences but instead to Y-specific repetitive sequences, often dispersed to multiple sites on the Y chromosome. (Technically, the term STS is reserved for single-copy sequence loci (10), but here we use the term more broadly to include these other PCR-defined loci.) An STS was recognized as likely to be repetitive if (i) it was present in numbers of YAC's far exceeding the expected 4.5-fold redundancy of the library; (ii) deletion mapping of the STS revealed multiple Y loci (11); (iii) it was derived from a clone shown to contain Y-specific repeats (by hybridization to Southern blots of human genomic DNA); or (iv) in a few cases, it appeared to join YAC contigs that were otherwise clearly separate.

YAC's containing Y-specific repetitive STS's could be placed only if they also contained or could be linked to single-copy STS's. For example, 18 YAC's were found to contain the three repetitive STS's sY60, sY61, and sY62. Some of these also contained single-copy STS's, allowing the

sY60-sY61-sY62 repeats to be mapped to two separate positions on Yp (Fig. 2, intervals 3C and 4A). In certain regions, particularly interval 6, many YAC's contained large numbers of Y-specific repetitive STS's (and, as a result, in Fig. 2 the lengths of these YAC's appear exaggerated); these STS's may derive from repeat arrays of considerable complexity. Although 31 YAC's containing only repetitive STS's were omitted, most YAC's containing Y-specific repetitive STS's could be positioned on the map. A useful by-product of the Y-specific repetitive nature of these STS's was a substantial increase in the number of Y-chromosomal DNA loci mapped (207 in all), ultimately improving the resolution of the resulting map.

Closing gaps. After assembly of contigs by STS content was complete, nine gaps remained. Because the STS's had been roughly ordered by deletion mapping (11), most of the contigs were ordered and oriented along the chromosome, which simplified the closing of gaps. After STS content mapping, gaps exist either because adjacent contigs fail to overlap (real gaps) or because the contigs actually do overlap but the region they share contains no STS (undetected overlaps). Real gaps must be closed by isolating additional YAC's, typically by screening libraries for STS's generated from the ends of YAC's flanking the gap (21). Undetected overlaps need only be revealed, either by finding an STS common

Fig. 1. Assembly of YAC contigs and ordering of STS's by STS content mapping. **(A)** Results of testing 11 Y-derived YAC's for the presence or absence of 18 STSs. The STS's present (indicated by a plus sign) in each YAC were initially ascertained during screening of the YAC library; all were confirmed by repeating PCR assays on DNA's from single colonies. Some negative results from library screening (those indicated by a minus sign) were reproduced on single colonies; no attempt was made to reproduce most negative results from library screening (indicated by the absence of any symbol). **(B)** Inferred overlaps among YAC's and inferred order of STSs, that is, the most economical interpretation of the data in (A). STS's stacked in piles are not ordered with respect to each other. Each YAC is represented by a black bar whose length is proportional to the number of STS's spanned; the drawing is otherwise not to scale. Some of the data in (A) (that is, the absence of sY41 and sY42 in yOX110, the absence of sY42 in yOX135) is inconsistent with the most likely order of STS's on the Y chromosome. These inconsistent data may reflect deletions within the inserts of yOX110 and yOX135, as represented by narrowed black bars.
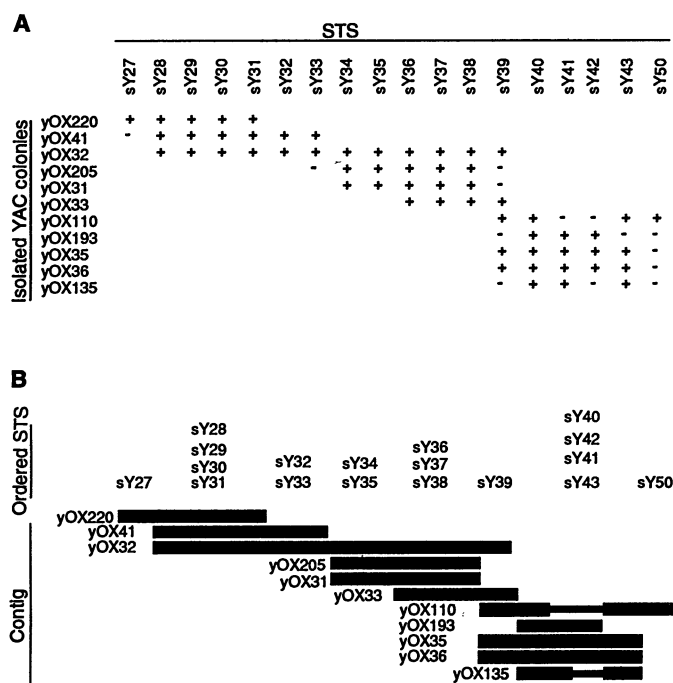

Fig. 2 (next page). Physical map of the euchromatic region of the Y chromosome. The figure is organized as two panels, one above the other. STS's (and three sites of *DYS7* hybridization; 207 loci in all) are listed near the bottom of each panel, along with the locations of eight genes and pseudogenes. The STS content of 196 overlapping YAC's is shown above (as in Fig. 1B). A single contig extends from sY4 [àn STS within 400 kb of the Yp telomere (23)] across the centromere to sY159 (within heterochromatic repeats comprising most of distal Yq). The two panels are joined through yOX186. Four overlaps resting solely on fingerprint data are indicated by black knobs at YAC ends. The pseudoautosomal region and deletion map intervals (11) are indicated, as are subintervals defined by YAC's. Within a ˘subinterval, STS's are not ordered with respect to each other. Subintervals 3C.4 and 6E.3/6F.1̄ comprise complex repetitive arrays in which STS's cannot be ordered; omitted here are the narrowed black bars elsewhere depicting internal YAC deletions. Regions known to contain X-homologous sequences, Y-specific repeats, alphoid repeats (and the centromere), satellite-3 repeats, or *DYZ1/DYZ2* heterochromatic repeats are indicated in color. The PCR conditions for sY183 are standard (11); the oligonucleotides are 5'-ACAGCATTACCTTGCCTCTG-3' and 5'-ATTCAATTGAACACTTGGTGTATG-3' and give a 152-bp product.

Heterochromatic
DY21/DYZ2 Repeats

Alphoid
Repeats

Satellite 3
Repeats
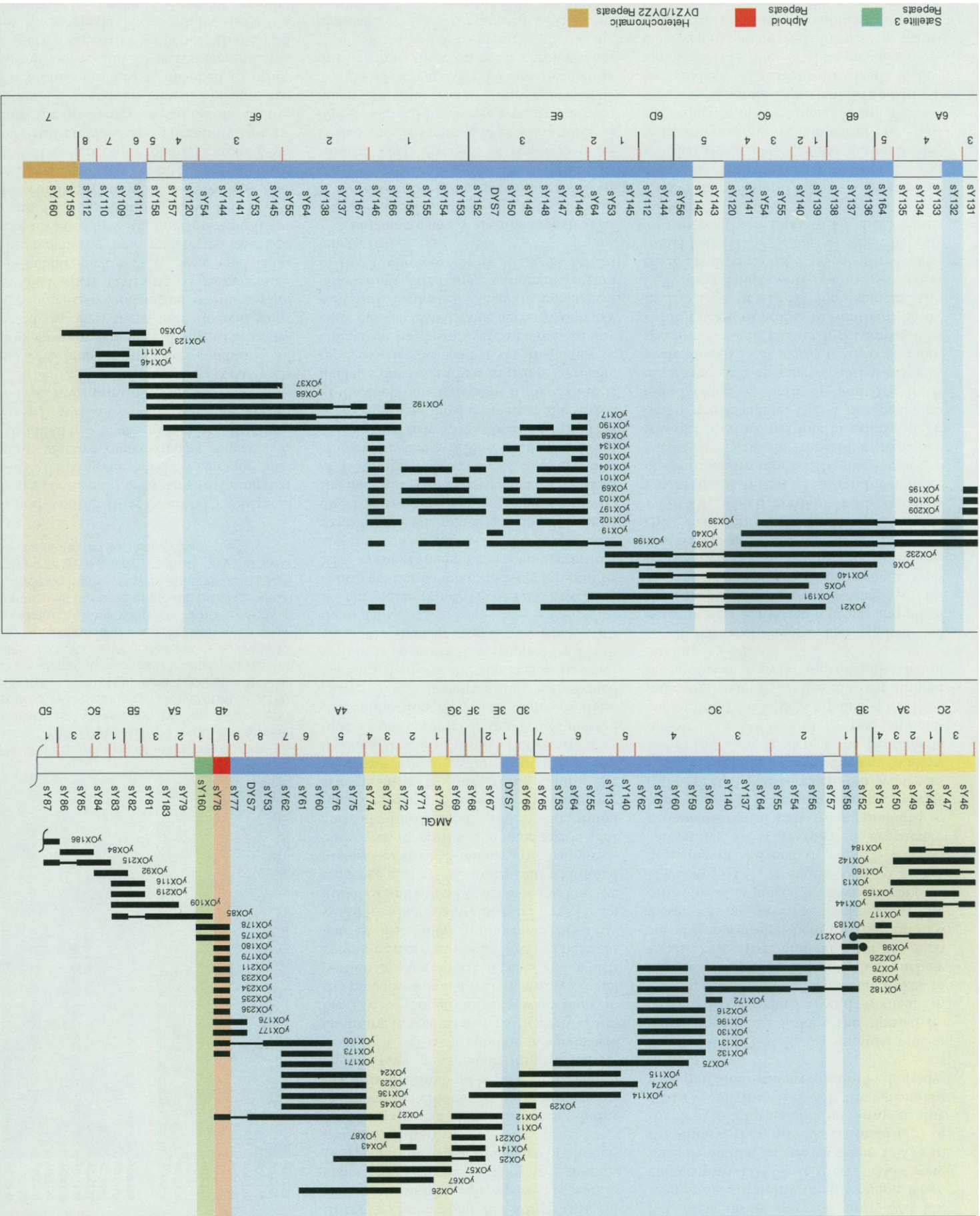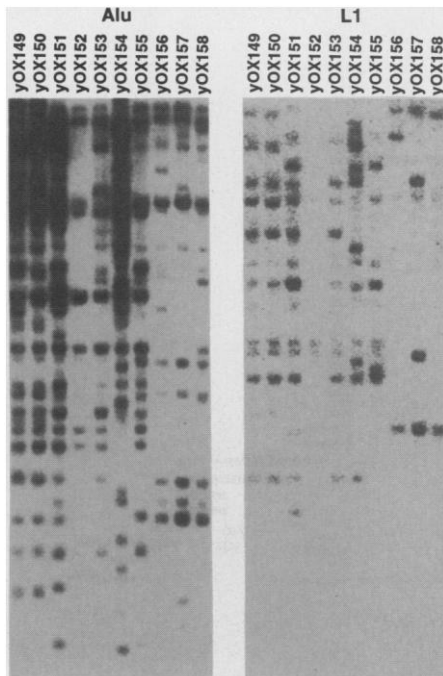
Fig. 3. Fingerprinting of restriction enzyme–digested YAC DNA's with common interspersed repeats as hybridization probes. DNA's from isolated YAC-bearing clones (yOX149, yOX150, yOX151, . . . yOX158) were digested with Hind III, transferred to nylon membranes, probed with either two Alu oligonucleotides or a single L1 oligonucleotide, and visualized with a Fuji phosphoimager. Overlapping YAC's have bands in common. In the case of the Alu probing, where hybridizing fragments were abundant and larger fragments were poorly resolved, only the lower portion of the gel was scrutinized.

to both contigs (that is, an STS generated from a contig end) or by demonstrating that YAC's flanking the gap show common fragments when a human repeat sequence is hybridized to a Southern blot of restriction-digested YAC DNA's, a technique known as YAC fingerprinting (22).

When STS generation from YAC ends and YAC fingerprinting were combined, all nine apparent gaps were shown to be undetected overlaps; there were no real gaps. Five previously undetected overlaps were revealed when YAC-end STS's generated by "bubble-anchor PCR" were used (21). The remaining four undetected overlaps were revealed by YAC fingerprinting (Figs. 2 and 3). Fingerprinting of all 196 YAC's also confirmed most of the overlaps that had been detected by STS content mapping. The generation of STS's from YAC ends also provided an opportunity to estimate the frequency of chimerism in the library. Of 47 ends examined, 24 mapped to the Y chromosome, 10 mapped to other chromosomes, and 13 were common, interspersed repetitive sequences that could not be mapped. These results suggest a chimerism frequency of about 59 percent.

[If the Y chromosome has a high concentration of common, interspersed repeats as compared with the rest of the genome (11), then the actual frequency of chimerism may be lower.]

Two principal difficulties arose during the pursuit of closure: (i) YAC fingerprinting with repetitive probes was not useful in regions poor in common interspersed repeats and (ii) STS content mapping was confusing in regions with particularly unstable YAC's. These problems were acute in the pseudoautosomal region, whose central portion is poor in L1 repeats and which contained many unstable YAC's. The problem of instability is illustrated by YAC yOX44, which contained STS sY11 on initial examination but appeared to lose the sequence as the YAC underwent additional passages in yeast. Similarly, sY7 had previously been mapped to the same region (23) and identified three YAC's on initial screening, but the sequence was ultimately not retained by any YAC. Thus, the pseudoautosomal region contains a few linkages that are historical, that is, after linkage was confirmed, the YAC's deleted the connecting STS's. In some of these cases, YAC fingerprinting reconfirmed overlaps. Fortunately, this region has previously been restriction mapped by PFGE (23, 24), and our map is consistent with these data.

The extent of overlap of YAC's within a contig cannot be quantitatively assessed by STS content mapping or fingerprinting, and thus the length of the 196-YAC contig spanning the euchromatic region cannot be directly determined at this time. However, the map's size can be estimated on the basis of YAC size, redundancy of coverage, and chimera frequency (25). We calculated that the physical map spans about 28 Mb. This is probably a minimal estimate of the length of the euchromatic region, if there is, as is likely, a downward bias in sizing Y-specific repetitive regions and if the frequency of chimerism has been overestimated. In any case, our calculated value agrees reasonably well with estimates, based on cytological observations, that the euchromatic region of the Y chromosome is 30 to 40 Mb in length (5).

Centromere and Y-specific repeats. Deletion interval 4B (Fig. 2), the only segment present on all independently segregating Y chromosomes, must contain the centromere (11). Analysis of this region revealed that it contains a block of alphoid repeats flanked by diverse repetitive arrays, much like other human chromosomes. The Y alphoid repeats (DYZ3) are sufficiently distinct from those on other chromosomes that it was possible to identify a Y-specific alphoid STS (sY78). Fourteen YAC's containing sY78 were isolated from the library,

but their mean size was only 270 kb, suggesting instability of this region when cloned into YAC's. If these YAC's were initially similar in size to other YAC's in the library, then a crude estimate (25) of the length of the alphoid array in the XYYYY cell line is 1 Mb, in agreement with previous estimates from PFGE analysis (26).

Of the 14 YAC's, 6 contained one or more other STS's linking the alphoid repeats to both proximal Yp and proximal Yq. On Yp, the alphoid sequences are linked to complex Y-specific repeats that include sY53, sY60, sY61, and sY62, all of which are also found elsewhere on the chromosome. On Yq, the STS closest to the centromere, as judged by deletion mapping (11), was sY81, a single-copy STS. Sequencing of the end of yOX85, a YAC containing sY81, revealed a degenerate pentameric repeat with striking similarity to both the DYZ1 heterochromatic repeats [(27), sY160] and to satellite-3 sequences (28). The similarity was sufficient to allow the sY160 assay to detect the pentameric repeat in yOX85—and in two alphoid-containing YAC's, yOX175 and yOX178—indicating that sY81 is linked to the alphoid repeats via an array of sequences like that of satellite-3 (29).

Of the euchromatic loci studied, 40 percent were Y-specific repeats, and all of these mapped to Yq and proximal Yp (Fig. 2). The distal portion of euchromatic Yq is represented almost exclusively by repetitive STS's, and these repeats are linked to the heterochromatin via one YAC (yOX50), which contains both the heterochromatic repeat sY159 and a number of euchromatic repeats. In many cases, a particular Y-specific repetitive sequence was dispersed to multiple locations on Yp or Yq. For example, the sY55 repeats are found twice on both Yp and Yq. Such dispersion of Y-specific repeats was found even though PCR assays tend to be more specific and detect fewer loci compared to hybridization of probes to Southern blots. In the case of sY132, for example, the PCR assay detects only one locus whereas the PCR product itself reveals three additional loci (DYS7 in Fig. 2) when hybridized to Southern blots of all the Y chromosomal YAC's.

X-Y homology. Through a mixture of common ancestry, transpositions, and translocations that occurred during evolution, and ongoing recombination, much of the sequence of the human X and Y chromosomes is similar. Of the euchromatic Y loci studied, 25 percent are clearly X homologous (11) (Fig. 2). The pseudoautosomal region, where X-Y identity is maintained by frequent recombination during male meiosis, occupies the most distal 2.7

64

Mb (23, 24) of both Yp and Xp. The pseudoautosomal region carries genes encoding a GM-CSF (granulocyte-macrophage colony-stimulating factor) receptor subunit (CSF2RA) (30) and a glycoprotein involved in T cell adhesion (MIC2) (31).

Proximal to the pseudoautosomal region on Yp is a comparably sized region homologous to Xq21. These Xq21-homologous sequences are present on Yp because of their transposition from the X chromosome during recent human evolution (32). These sequences had been shown to exist on Yp as two blocks (33), which appear on the YAC contig map as a large distal segment of approximately 3 Mb (sY20 through sY52) and a much smaller proximal segment [sY73 (DXYS1) and sY74]. Most of the PCR assays derived from these Xq21-homologous sequences could not distinguish between X- and Y-derived YAC's, so hybridization probes detecting chromosome-specific restriction fragments were used to exclude seven X-chromosomal YAC's isolated in the screen. Close to but distinct from the small proximal segment of Xq21 homology is the amelogenin gene (AMGL), which has a closely related homolog in Xp22 (34).

Between the pseudoautosomal and Xq21-homologous regions is an intensively studied segment of the Y chromosome. In addition to the sex-determining gene SRY (35), this 280-kb region contains two genes with X homologs, the ribosomal protein gene RPS4Y (36) and the zinc finger gene ZFY (37). In the case of RPS4Y and ZFY, X-Y nucleotide similarity is essentially restricted to coding sequences and does not extend to introns or flanking sequences (38).

Yq also contains regions of X homology. A concentration of Xp22-homologous sequences is found in intervals 5E-5I, on proximal Yq. These X-homologous sequences include the complex CRI-S232 repeats (sY91) (39) and two nonprocessed pseudogenes, one related to the X-linked steroid sulfatase gene (STSP) (40) and the other related to the Kallman gene (KAL-Y) (41). Homologies near the distal extremes of Yq and Xq have been reported (42), but we have not identified corresponding YAC's, perhaps because our YAC cloning was biased against telomeric regions.

Completeness, accuracy, and resolution of the map. Our results allow us to address the question of the completeness of the coverage of the Y chromosome. The map comprises one contig that spans virtually the entire euchromatic Y chromosome. This alone indicates that a large portion of the chromosome has been cloned. Nonetheless, there may still be small segments of the chromosome that are not represented in the overlapping array of recombinant DNA clones. The absence of such segments, fall-

ing between STS's, could result from deletion of unstable regions in YAC's or from the existence of large blocks of repetitive DNA not spanned by single YAC's. One indication of the completeness of the map is the fact that all but four of the 160 Y chromosomal STS's are represented. Two of these were immediately subtelomeric and thus were probably absent from the library because of its method of construction. This leaves two unrepresented STS's, or 1.2 percent of the total, suggesting that more than 98 percent of the euchromatic Y chromosome is present.

There is no simple quantitative measure of the reliability of this map. However, there are three indications that in most cases the YAC's and STS's have been ordered correctly by STS content mapping. First, in most regions, assignments of order were based on redundant independent observations on multiple YAC's. Inconsistencies were detected in about 5 percent of such situations. Most inconsistencies were likely the result of YAC's deleting unstable sequences that contained STS's. Second, virtually all YAC overlaps detected by STS content mapping were confirmed by YAC fingerprinting. Third, no inconsistencies were detected between the STS order derived by deletion mapping of human individuals (11) and that derived by STS content mapping of YAC's (recognizing, however, that these two STS ordering exercises were not completely independent). If errors in ordering do exist, they are most likely to be found in regions containing Y-specific repetitive sequences.

The degree of resolution is one of the chief determinants of the utility of a map. In this physical map, 207 DNA loci were assigned to 127 ordered segments, which represents an average spacing of about 140 kb between loci and about 220 kb between ordered loci. This set of ordered STS's should provide a useful framework for analysis of the Y chromosome and could be used to rapidly regenerate the physical map from any improved large-insert clone library (that is, one in which chimerism and deletion are uncommon).

Construction of complete physical maps of human chromosomes should facilitate positional cloning of genes for a given phenotype, expedite the more comprehensive task of identifying all genes, and provide substrates for large-scale sequencing. This physical map of the Y chromosome should be useful in examining Y chromosomal polymorphism among human subpopulations, in studying the comparative anatomy of the chromosome in primates, and in fine mapping of X-Y homologous sequences in an effort to elucidate the evolution and functional relationship of the sex chromosomes.

Applicability of strategy to other chromosomes. Our method of physical mapping is readily applicable to other human chromosomes, provided that a large number of STS's can be generated from the chromosome. For the Y chromosome, the STS's were first subjected to deletion mapping of DNA samples from patients carrying partial Y chromosomes (11). This prior information made it possible to assemble YAC contigs without resorting to any sophisticated computation and also resulted in immediate ordering and orienting of contigs, which facilitated the closure of gaps. However, contig assembly would still be possible in the absence of such information—as demonstrated by the ease with which contigs of the large Xq21-homologous region were assembled even though only half of the STS's in this region were mapped. It would then be necessary to order and orient the resulting contigs along the chromosome, perhaps by fluorescent hybridization in situ with terminal YAC's as probes or by incorporating previously ordered markers into the contigs as anchor points. For any chromosome but the Y, it might be possible to simultaneously assemble, order, and orient contigs with a collection of STS's in which some were genetically ordered polymorphisms. In our analysis, the deletion mapping information was particularly useful in overcoming difficulties posed by Y-specific repeats occurring in large, dispersed blocks. Such repeats will probably be encountered less frequently on other human chromosomes.

The demonstration that a physical map of the Y chromosome can be efficiently constructed from a total human YAC library suggests that similar maps can be made for other chromosomes without the laborious step of building chromosome-specific YAC libraries. (In the physical mapping exercise described here, the most arduous steps included the construction of the YAC library and the sorting out of ambiguities posed by Y-specific repeats; the screening of the library was relatively straightforward.) Although there may be a higher rate of chimerism in whole genomic libraries than in chromosome-specific libraries, this does not appear to pose a serious problem for STS content mapping. In conclusion, the strategy should be widely applicable to other human chromosomes and is also a logical choice in any genome wide-mapping effort.

REFERENCES AND NOTES

1. A. Coulson, J. Sulston, S. Brenner, J. Karn, Proc. Natl. Acad. Sci. U.S.A. **83**, 7821 (1986); M. V. Olson et al., ibid., p. 7826; Y. Kohara, K. Akiyama, K. Isono, Cell **50**, 495 (1987).
2. D. T. Burke, G. F. Carle, M. V. Olson, Science **236**, 806 (1987).
3. D. Garza, J. W. Ajioka, D. T. Burke, D. L. Hartl,

ibid. **246**, 641 (1989); A. Kuspa, D. Vollrath, Y. Cheng, D. Kaiser, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8917 (1989); A. Coulson, R. Waterston, J. Kiff, J. Sulston, Y. Kohara, *Nature* **335**, 184 (1990); J. Ajioka *et al.*, *Chromosoma* **100**, 495 (1991); E. Maier *et al.*, *Nat. Genet.* **1**, 273 (1992).

4. F. E. Abidi, M. Wada, R. D. Little, D. Schlessinger, *Genomics* **7**, 363 (1990); D. Schlessinger *et al.*, *ibid.* **11**, 783 (1991); R. D. Little, G. Pilia, S. Johnson, M. D'Urso, D. Schlessinger, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 177 (1992).
5. N. E. Morton, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7474 (1991).
6. A. A. Sandberg, Ed., *The Y Chromosome* (Liss, New York, 1985).
7. M. Frommer, J. Prosser, P. C. Vincent, *Nucleic Acids Res.* **12**, 2887 (1984); K. D. Smith, K. E. Young, C. C. Talbot, Jr., B. J. Schmeckpeper, *Development* **101** (suppl.), 77 (1987).
8. E. D. Green and M. V. Olson, *Science* **250**, 94 (1990); E. D. Green and P. Green, *PCR Methods Appl.* **1**, 77 (1991).
9. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985); R. K. Saiki *et al.*, *ibid.* **239**, 487 (1988).
10. M. Olson, L. Hood, C. Cantor, D. Botstein, *ibid.* **245**, 1434 (1989).
11. D. Vollrath *et al.*, *ibid.* **258**, 52 (1992).
12. E. D. Green, H. C. Riethman, J. E. Dutchik, M. V. Olson, *Genomics* **11**, 658 (1991).
13. R. Arratia, E. S. Lander, S. Tavaré, M. Waterman, *ibid.*, p. 806.
14. To construct the YAC library, we used a protocol adapted from that of H. M. Albertsen *et al.* [*Proc. Natl. Acad. Sci. U.S.A.* **87**, 4256 (1990)]. In brief, OXEN lymphoblastoid DNA was purified in agarose plugs, partially digested with Eco RI (digestion was limited by the presence of Eco RI methylase; D. R. Smith, A. P. Smyth, D. T. Moir, *Methods Enzymol.*, in press) and size-fractionated by pulsed field gel electrophoresis. Fragments longer than 600 kb were isolated, ligated to phosphorylated pYAC4 arms, again size-selected to eliminate unligated vector, transfected into AB1380 spheroplasts, and plated onto uracil-deficient medium. Colonies were selected and patched onto AHC medium (deficient in uracil and tryptophan) and then into AHC-containing 96-well plates for DNA preparation and freezing.
15. L. Sirota, Y. Zlotogora, F. Shabtai, I. Halbrecht, E. Elian, *Clin. Genet.* **19**, 87 (1981).
16. All screening was done with PCR, the products being detected with agarose gel electrophoresis and ethidium bromide staining. Pooled DNA's were used as templates. Each pool represented six 96-well plates (576 YAC's), arranged as a three-plate by two-plate array with 24 rows and 24 columns (step 1). Each pool was divided into 12 subpools, each representing four rows or four columns (step 2). For positive subpools, the eight component rows and columns were individually tested (step 3), resulting in identification of a single positive well. Positive results were confirmed with DNA's from isolated colonies. In many cases, step 2 screening suggested that multiple STS's were present in the same YAC. Step 3 screening was then restricted to one STS; the presence of other STS's was subsequently confirmed on isolated colony DNA. In cases where an STS was present in two, three, or more YAC's in a single pool, step 3 screening became increasingly cumbersome.
17. G. Chu, D. Vollrath, R. W. Davis, *Science* **234**, 1582 (1986).
18. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
19. D. L. Neil *et al.*, *Nucleic Acids Res.* **18**, 1421 (1990).
20. H. J. Cooke, W. R. A. Brown, G. A. Rappold, *Nature* **317**, 687 (1985); D. C. Page *et al.*, *Genomics* **1**, 243 (1987).
21. Sequence from YAC ends was obtained with "vectorette" or "anchor-bubble" PCR [J. Riley *et al.*, *Nucleic Acids Res.* **18**, 2887 (1990)]. The following "bubble" adaptor was designed to be compatible with the M13 (−21) sequencing primer and to allow automated sequencing of PCR products isolated from low melting agarose

$$5'-A(AGCT)TCCGGTACATGATCGAGGGGACTGAC-$$
$$AACGAACGAACGGTTGAGAAGGGAGAG-3'$$

$$3'-GGCCATGTACTAGCTCCTGACCGGCAGCAAAA-$$
$$TGGCGATCCTCTTCCCTCTC-5'$$

After ligation of the adaptor to yeast DNA digested with Hinf I or Rsa I, ends were amplified by hemi-nested PCR with the use of the bubble primer 5'-TAGCGGTAAAACGACGGCCA-3' in conjunction with primers based on the pYAC4 vector, as described (E. D. Green, *Methods Mol. Genet.*, in press). Sequencing was performed and PCR primers were chosen as described (*11*).
22. In our case, YAC fingerprints were examined on Southern blots of Eco RI– or Hind III–digested DNA's probed with either a degenerate L1 oligonucleotide (5'-TGGGTGCAGC(AG)CACCA(AG)C-ATGGCACATGGTATACATATGTAAC(AT)AACCT-GCAC-3') or a pair of degenerate Alu oligonucleotides—(5'-TGAGC(CT)(GA)(AT)GAT(CT)(GA)(CT)(GA)CCA(CT)TGCACTCCAGCCTGGG-3' and 5'-GCCTCCCAAAGTGCTGGGATTACAG-G(CT)(GA)TGAGCCA-3')—[D. L. Nelson, *Methods* **2**, 60 (1990)]. Overlap was confirmed when YAC's had two or more hybridizing fragments in common.
23. G. Rappold and H. Lehrach, *Nucleic Acids Res.* **16**, 5631 (1988); A. Henke *et al.*, *Am. J. Hum. Genet.* **49**, 811 (1991).
24. C. Petit, J. Levilliers, J. Weissenbach, *EMBO J.* **7**, 2369 (1988); W. R. A. Brown, *ibid.*, p. 2377.
25. Chromosome length *L* can be estimated by

$$L = l(n/r)k$$

where *l* is the mean YAC length, *n* is the total number of Y chromosomal YAC's, *r* is the redundancy of coverage, and *k* is a correction factor for chimerism. Redundancy *r* can be estimated by

$$r = a/b$$

where *a* is the sum of the number of STS's on each YAC and *b* is the number of STS's on the chromosome. Chimerism factor *k* is the fraction of YAC ends found to be of Y origin

$$k = c/d$$

where *c* is the number of Y chromosomal ends and *d* is the total number of ends mapped. To calculate *L*, we assumed *l* = 650 kb (the average length of YAC's in the library, recognizing that instability had shortened many of the Y chromosomal YAC's), *n* = 254 (including approximately 20 YAC's identified in the initial screen but not isolated), *a* = 845, *b* = 207, *c* = 24, and *d* = 34.
26. R. Oakey and C. Tyler-Smith, *Genomics* **7**, 325 (1990).
27. Y. Nakahori, K. Mitani, M. Yamada, Y. Nakagome, *Nucleic Acids Res.* **14**, 7569 (1986).
28. H. J. Cooke and J. Hindley, *ibid.* **6**, 3177 (1979); P. L. Deininger, D. J. Jolly, C. M. Rubin, T. Friedmann, C. W. Schmid, *J. Mol. Biol.* **151**, 17 (1981).
29. An STS (sY79) generated from satellite-3–like sequences at the end of yOX85 detected no other YAC's.
30. N. M. Gough *et al.*, *Nature* **345**, 734 (1990).
31. P. J. Goodfellow, S. M. Darling, N. S. Thomas, P. N. Goodfellow, *Science* **234**, 740 (1986); C. Gelin *et al.*, *EMBO J.* **8**, 3253 (1989).
32. D. C. Page, M. E. Harper, J. Love, D. Botstein, *Nature* **311**, 119 (1984).
33. G. Vergnaud *et al.*, *Am. J. Hum. Genet.* **38**, 109 (1986).
34. E. C. Lau, T. K. Mohandas, L. J. Shapiro, H. C. Slavkin, M. L. Snead, *Genomics* **4**, 162 (1989); Y. Nakahori, O. Takenaka, Y. Nakagome, *ibid.* **9**, 264 (1991).
35. A. H. Sinclair *et al.*, *Nature* **346**, 240 (1990).
36. E. M. C. Fisher *et al.*, *Cell* **63**, 1205 (1990).
37. D. C. Page *et al.*, *ibid.* **51**, 1091 (1987).
38. A. Schneider-Gadicke, P. Beer-Romero, L. G. Brown, R. Nussbaum, D. C. Page, *ibid.* **57**, 1247 (1989).
39. R. G. Knowlton, C. A. Nelson, V. A. Brown, D. C. Page, H. Donis-Keller, *Nucleic Acids Res.* **17**, 423 (1989).
40. P. H. Yen *et al.*, *Cell* **55**, 1123 (1988).
41. S. Guioli *et al.*, *Nature Genet.* **1**, 337 (1992); B. Incerti *et al.*, *ibid.*, in press.
42. H. J. Cooke, W. A. R. Brown, G. A. Rappold, *Nature* **311**, 259 (1984); A. Pedicini, G. Camerino, R. Avarello, S. Guioli, O. Zuffardi, *Genomics* **11**, 482 (1991).
43. We thank L. Brown, T. Dorman, P. Green, D. Housman, K. Kusumi, E. Lander, D. Moir, M. Olson, J. Segre, J. S. Smith, D. Smith, V. Stanton, and M. Velez-Stringer for technical assistance, helpful discussions, and advice. Supported by grants from the National Institutes of Health and the Searle Scholars Program/Chicago Community Trust, a Markey Foundation fellowship (S.F.), and a Smith Kline Beecham fellowship of the Life Sciences Research Foundation (D.V.).

20 July 1992; accepted 20 August 1992