J Mol Evol (1991) 33:133–141

Journal of Molecular Evolution


# Evolution and Sequence Analysis of a Human Y-Chromosomal DNA Fragment

E.C. Whisenant,[1] B.K.A. Rasheed,[1] H. Ostrer,[2] and Y.M. Bhatnagar[1]*

[1] Departments of Structural & Cellular Biology, and Medical Genetics, University of South Alabama, College of Medicine, Mobile, AL 36688, USA
[2] Division of Genetics, Department of Pediatrics, University of Florida, College of Medicine, Gainesville, FL 32610, USA

**Summary.** A Y-chromosomal DNA fragment has been isolated from a human Y-Charon 21A recombinant library. Evolutionary analysis of 1F5 indicates that the size and sequence of this fragment have been conserved in higher primates. Deletion mapping and in situ hybridization analysis have localized 1F5 to the middle euchromatic portion of the long arm of the human Y chromosome at Yq11.2. Sequence analysis revealed the presence of an atypical Alu element and two regions rich in polypyrimidine–polypurine residues.

**Key words:** Y-chromosome — DNA — Human — Primate — Evolution — PUPPY sequence — Alu element

## Introduction

The human Y chromosome is one of the smallest chromosomes and its presence is critical for male sexual determination and differentiation (Stewart 1983). Its short arm (Yp) is rather minute and contains at least one genetic locus important in male sex determination (Bishop et al. 1985). The long arm (Yq) consists predominantly of heterochromatin, which accounts for approximately 50–70% of the total DNA present in this arm (Kunkel et al. 1979). Available evidence indicates that the heterochromatic region (Yq12) consists of at least two major repeated DNA families (Cooke and McKay 1978). Surprisingly, deletions in the heterochromatic region of Yq or its translocation to other chromosomes in the female genome do not appear to lead to any detectable abnormalities (Buhler 1985). Extensive polymorphism of the Yq12 region of the Y chromosome in the human population also suggests the absence of functional genetic loci in this segment (Bobrow et al. 1971).

Available molecular and genetic data indicate a relative deficiency of functional genetic loci on the human Y chromosome (Weissenbach et al. 1989). The primary locus regulating sex determination (SRY) has been assigned to the short arm (Yp11.2-Ypter) of this chromosome (Vergnaud et al. 1986; Palmer et al. 1989; Sinclair et al. 1990). Other loci assigned to the Y chromosome include genes for the H-Y antigen, ZFY, MIC2 antigen, GM-CSF receptor, JA36, amelogenin, growth and stature, hypertrichosis, and spermatogenesis. In addition, a number of pseudogenes, presumably formed by the retroposition of autosomal or X-chromosomal loci, have also been assigned to this chromosome (Weissenbach et al. 1989).

A large number of Y chromosomal DNA fragments have been isolated, however, only a few have been studied in any detail. When single and low-copy probes have been isolated at random from flow sorted Y-DNA libraries, they have fallen into the following categories: 35% were exclusively Y linked, 46% showed X–Y shared sequences, 15% were Y–autosome shared, and 4% were X–Y–autosome shared (Bishop et al. 1984; Affara et al. 1986). Although much is known about the Y-linked repeated elements (Willard 1985), sequences unique to this chromosome remain less well defined. A comprehensive understanding of the evolution and genomic

*Offprint requests to:* Y.M. Bhatnagar

organization of the human Y chromosome will require a detailed knowledge of the single and low-copy sequences unique to this chromosome.

We have been interested in studying human Y chromosome-specific fragments that may be conserved during primate evolution. Our strategy for the isolation of such fragments has included the screening of a human Y chromosome library with total human male and female DNA probes to select clones consisting of male-specific fragments. These clones were further screened with total male mouse DNA to select conserved sequences. This approach has led to the isolation of two Y-linked clones designated λYH1F5 (DYS128, hereafter abbreviated 1F5) and λYH2D6 (DYS129), which are conserved in higher primates.

The studies described here provide evidence for the localization of 1F5 to the middle euchromatic portion of the long arm of the Y chromosome, and for its ancient association with the primate Y chromosome. Sequence analysis of 1F5 has revealed the presence of an atypical Alu element, and two regions with striking strand asymmetry typical of DNA involved in the formation of unusual non-B type of structures. Such sequences have not been reported to be present in the human Y chromosome until now, and may have important biological consequences in terms of their structure and function.

## Materials and Methods

*Materials.* The λCharon 21A human Y chromosome library (LAOYNS01) used in this study was obtained from the American Type Culture Collection, Rockville, MD. The pTZ series of phagemid vectors, and sequencing reagents were from U.S. Biochemical Corp. (Cleveland, OH). The random-primer labeling kit was acquired from Pharmacia Biotechnology (Piscataway, NJ). The [$\alpha$-$^{32}$P]dCTP (>3000 Ci/mmol) and the [$\alpha$-$^{35}$S]dATP (>1000 Ci/mmol) were supplied by Amersham Corp. (Arlington Heights, IL). For the nick-translation reaction, $^{3}$H-dCTP (40–60 Ci/mmol), $^{3}$H-TTP (90–110 Ci/mmol), and DNA polymerase I were purchased from New England Nuclear (Boston, MA); DNase I was a product of Worthington Biotechnologies (Freehold, NJ), and the dATP and dGTP were from Sigma Chemical Co. (St. Louis, MO). The various restriction enzymes were purchased from one of the following commercial suppliers: Brisco Ltd. (Winthrop, MA), Pharmacia Biotechnology, U.S. Biochemical Corp., or Bethesda Research Laboratories (Gaithersburg, MD). The fibroblastic cell lines: GM9996, GM2103, GM2626, GM2668, and GM2730, as well as the somatic cell hybrids GM06317 and GM06318B were obtained from the National Institute of General Medical Sciences (NIGMS) Human Genetic Mutant Cell Repository, Coriell Institute for Medical Research (Camden, NJ). The human lymphoblastic cell lines, Oxen (49, XYYYY) and VGD1 were gifts from David Page and V.G. Dev, respectively. Primate blood samples used for the isolation of high molecular weight DNA were obtained from the Yerkes Primate Center (Atlanta, GA).

*DNA Isolation, Southern Blotting, and Hybridization.* High molecular weight DNA was prepared from leukocytes as described (Davis et al. 1986). Restriction enzyme digestions, electrophoresis, and dot and Southern blotting were performed essentially as described by Davis et al. (1986). Hybridizations were performed following the method outlined by Maniatis et al. (1982). Radioactive probes labeled with $^{32}$P were prepared using the random-primer DNA labeling method of Feinberg and Vogelstein (1983). Following hybridization, blots were first washed three times in 2× SSC (standard saline citrate: 0.15 M NACl, 0.015 M sodium citrate)/0.1% SDS at 25°C for 15 min, and then at varying SSC concentrations depending on the desired stringency of washing. The conditions were as follows: low stringency (1–2× SSC/0.1% SDS), moderate stringency (0.5× SSC/0.1% SDS), or high stringency (0.1× SSC/0.1% SDS). The latter washings were performed twice at 65°C for 30 min. The filters were blotted dry and exposed to Kodak X-OMAT AR film for autoradiography.

We estimate that, under the experimental conditions, bands recognized at highest stringency washing conditions share at least 95% sequence similarity with the 1F5 probe. Bands identified at moderate stringency conditions share at least 85% sequence similarity with the 1F5 probe.

*DNA Sequencing.* The nucleotide sequence of both strands was determined by the dideoxy chain termination method (Sanger et al. 1977). The appropriate series of overlapping clones was generated by standard cloning techniques (Henikoff 1987). Occasionally, double-stranded sequencing was performed on fragments cloned in plasmid vectors pT7/T3-18 and pT7/T3-19 (Bethesda Research Laboratories). The deduced sequence was analyzed by the PCS DNA sequence analysis package (L.M. Lagramini and S.T. Brentano, University of Iowa, Iowa City, IA) and by Pustell Program, a DNA/protein sequence analysis package (Pustell and Kafatos 1984), from International Biotechnologies, Inc. (New Haven, CT). Two DNA sequence databases, GenBank and the European Molecular Biology Laboratory (EMBL), were searched for homologous sequences.

*In Situ Hybridizations.* Metaphase spreads of leukocyte chromosome preparations were prepared by standard methods as previously described (Yunis 1976). $^{3}$H-labeled probe was prepared by nick translation, and the hybridizations were performed as described (Lai et al. 1979). The radiolabeled 1F5 probe was used at a concentration of 50–200 ng/ml for in situ hybridization. Nuclear track emulsion autoradiography was performed by standard methods using Kodak type NTB2 (Harper and Saunders 1981; Harper et al. 1981). The number of grains per chromosome was determined by counting 50 metaphase spreads. Approximately 94% of the cells analyzed had 1–3 grains per mitosis.

*Cell Culture.* All cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM; Flow Laboratories, McLean, VA) according to Coriell Institute's recommendations. The DMEM was supplemented with 10% fetal bovine serum and antibiotic–antimycotic solution (10 mg/ml streptomycin, 10,000 units/ml penicillin, and 25 μg/ml amphotericin B; Sigma Chemical Co.). Cell cultures were maintained at 37°C in a 5% $CO_2$-enriched atmosphere.

## Results

### *Isolation of the Male-Specific Y-Chromosomal DNA Fragment 1F5*

Approximately, 1000 independent plaques from the λCharon 21A-Y library were screened with the two different probes using dot blots. The rationale for the approach used was as follows: The use of total

human female DNA probe allowed for the identification of those clones, which were exclusively Y-linked (at high stringency washing conditions). Only those clones negative (63%) in this assay were subjected to the second round of screening. The second probe used was total mouse male DNA. Approximately, 4–5% of these clones hybridized with this probe in the presence of excess, nonradioactive, female mouse DNA under low stringency washing conditions. This group of recombinants, therefore, consisted of sequences that may be partially conserved between mouse and human. Conservation, as defined here does not imply strict homology but simply the ability to hybridize with a heterologous probe under low to moderate stringency washing conditions. Evolutionary conservation may imply significance of the isolated DNA fragment as functional DNA is more likely to be conserved than nonfunctional DNA. These two screening steps led to the isolation of two Y-linked fragments, λYH1F5 and λYH2D6. Our studies on 2D6 are reported elsewhere (Rasheed et al. 1991).

*Size Analysis and Initial Characterization*

Agarose gel analysis indicated the size of 1F5 to be approximately 4.1 kb. This fragment was used as a probe to examine the Southern blots of EcoRI-restricted male and female genomic DNA in order to confirm the results seen in dot blots (Fig. 1). Additionally, the oxen lymphoblastic cell line (49, XYYYY) as well as the somatic cell hybrids GM06317 (retains a human Y) and GM06318B (retains a human X) were probed. As shown in Fig. 1B, under high stringency washing conditions, 1F5 recognizes a 4.1-kb EcoRI restriction fragment only in male DNA. The presence of a 4.1-kb fragment in the Y chromosome–somatic cell hybrid and an increased intensity of the 4.1-kb band in the oxen cell line further confirms the localization of the 1F5 fragment to the Y chromosome. Titration studies revealed that only 10–20 copies of 1F5 were present in the human male genome (data not shown). Thus, 1F5 represents a low-copy number sequence present on the human Y chromosome.

At moderate stringency washing conditions (Fig. 1A), a second male- and female-shared fragment of 6.7 kb was also detectable. Because the 6.7-kb fragment was not detected in either the X or Y somatic cell hybrids, this fragment is most likely autosomally located. Although the Southern blots show discrete length bands at high and moderate stringency washing conditions with the 1F5 probe, a smear is observed at low stringency washing conditions. These data suggest cross-hybridization of 1F5 to partially homologous DNA fragments, which are heterogeneous in size.

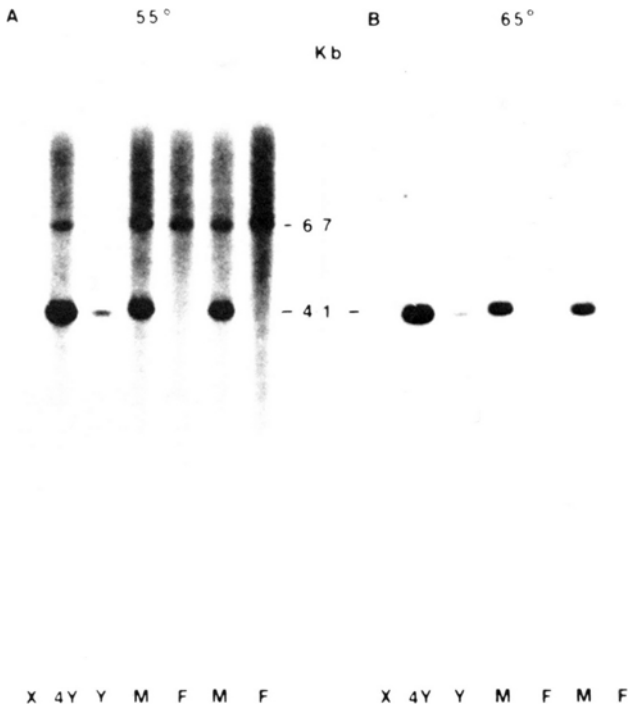


**Fig. 1.** Chromosomal assignment of 1F5. A Southern blot of human genomic DNA (10 μg) preparations cut with EcoRI, and probed with 1F5, is shown. Genomic DNA preparations were obtained from hamster/human X hybrid (X), oxen (XYYYY) cell line (4Y), hamster/human Y hybrid (Y), two human males (M), and two human females (F). Autoradiographs washed under moderate (**A**), and high (**B**) stringency washing conditions are shown here.

*Regional Localization of 1F5 by Deletion Mapping*

For regional mapping of the 1F5 fragment on the Y chromosome, the DNA preparations from a number of cell lines with known Y-chromosomal aberrations were examined. For these experiments, the DNA was digested with EcoRI, Southern blotted, and probed with 1F5. As seen in Fig. 2, a signal at 4.1 kb was apparent in the two X:Y translocations, GM2103 and VGD1 (lanes 4 and 6). In the other cell lines examined (GM9996, GM2626, GM2668, and GM2730), the 4.1-kb EcoRI fragment was not detected. From these data, it was apparent that 1F5 most likely maps to the euchromatic portion or the heterochromatic region of the Y chromosome (Yq11.2-Yq12). A summary of the karyotype of the cell lines surveyed and the results obtained are shown in Table 1.

*In Situ Hybridization*

The subchromosomal location of 1F5 was further refined by in situ hybridization. Analysis of the grain distribution in 50 cells (Fig. 3) indicates that 1F5 maps to the middle euchromatic portion (Yq11.2)

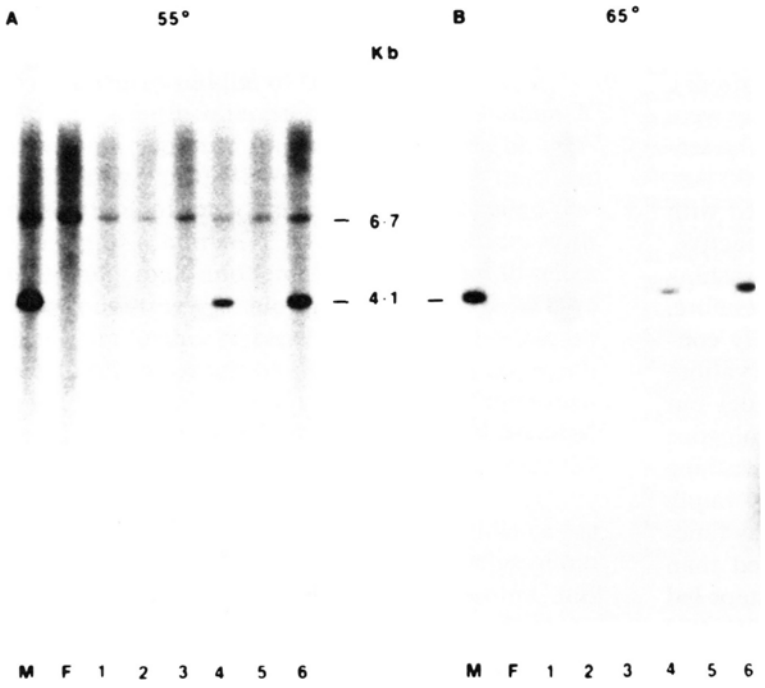


**Fig. 2.** Regional assignment of 1F5. Genomic DNA preparations (EcoRI-cut) from human male (M), human female (F), GM09996 (lane 1), GM2668 (lane 2), GM2730 (lane 3), GM2103 (lane 4), GM2626 (lane 5), and VGD1 (lane 6), respectively. Blots washed under moderate (**A**), and high (**B**) stringency conditions are shown.

**Table 1.** Deletion mapping analysis of 1F5

| Cell line | Sexual phenotype | Karyotype | 1F5[a] |
|---|---|---|---|
| GM2103 | F | 46,X,t(X;Y)<br>(Xpter > Xq11:Yq11 > Yqter) | + |
| GM2626 | M | 46,XX | − |
| GM2668 | Ambiguous | 46,X/46,S,del(Y)(pter > q12:) | − |
| GM2730 | Ambiguous | 45,X/46,X,del(Y)(pter > q11:) | − |
| GM9996 | M | 46,X,del(Y)(pter > q11:),inv(9)<br>(pter > p13::q13 > p13::q13 > qter) | − |
| VGD1 | F | 46,X,t(X;Y)<br>(Xpter > Xq11::Yq11 > Yqter) | + |

[a] The + or − symbols indicate the presence or absence of the 4.1-kb hybridizing band in Southern blots of EcoRI-digested DNA

of the Y chromosome. These results are consistent with the results obtained from the deletion mapping analysis.

### *Evolutionary Conservation*

In order to examine the evolutionary conservation of 1F5, a number of primate DNA samples were digested with EcoRI and subjected to Southern blot analysis. When probed with 1F5 and washed at high stringency conditions (Fig. 4), a single male-specific band was detected in human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), and orangutan (*Pongo pygmaeus*). Significantly, the size of the male-specific fragment detected by the 1F5 probe was well conserved in human, chimpanzee, and gorilla genomic DNA preparations. In orangutan, however, the male-specific band is larger (approx. 6.0 kb). No hybridization signal was detected in human, chimpanzee, gorilla, or orangutan females (Fig. 4), or in either sex of gibbon (*Hylobates lar*) or macaque (*Macaca mulatta*), under these washing conditions (data not shown).

When the Southern blots containing orangutan, gibbon, and macaque DNA were examined at moderate stringency washing conditions, other fragments of reduced sequence similarity with the 1F5 probe became apparent (Fig. 5). A male–female-shared 6.7-kb fragment, similar to the one seen in humans (Fig. 1), was observed in all species surveyed (including gorilla and chimpanzee; data not shown) except orangutan where it was smaller (6.2 kb; Fig. 5). In addition, male orangutan showed a 6.7-kb hybridizing band (Fig. 5). The hybridization pattern seen in macaque (Fig. 5) was similar to that of human, chimpanzee, and gorilla, although the male-specific 4.1-kb fragment was undetectable at high stringency conditions. By contrast, no male-specific bands were seen in gibbon at any of the washing conditions used (Fig. 5). We have repeated

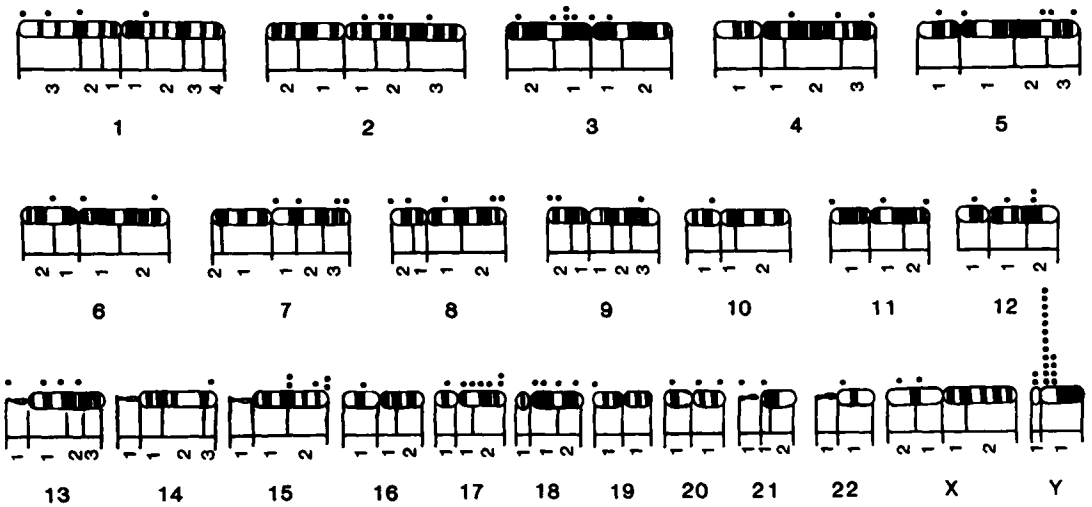


**Fig. 3.** Ideogram of 1F5 in situ hybridization. The distribution of grains on metaphase chromosomes is shown. The sequence was assigned to Yq11.2.

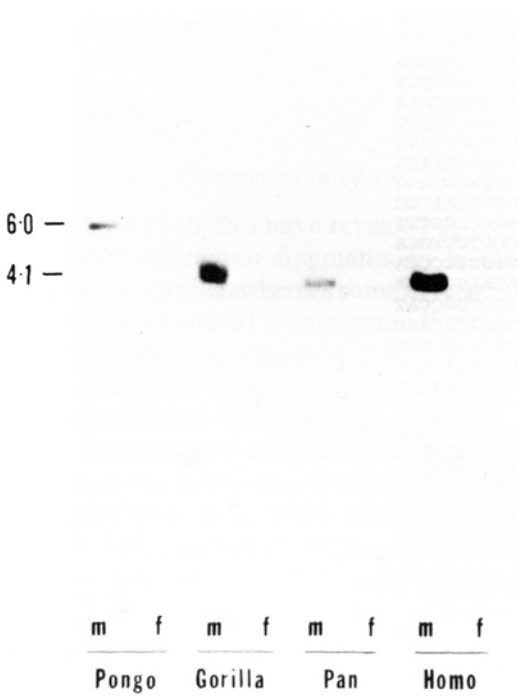


**Fig. 4.** Primate Southern blot at high stringency conditions. Male (M) and female (F) DNA (5 μg each) preparations from different primate species *Pongo* (orangutan), *Gorilla, Pan* (chimpanzee), and *Homo* (Human) were digested with EcoRI and resolved on a 0.7% agarose gel. A Southern blot of the gel was hybridized with the 1F5 probe. The blot was then washed at high stringency washing conditions (0.1× SSC/65°C) and autoradiographed.

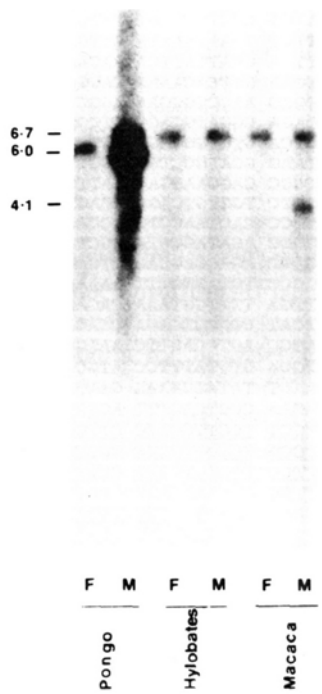


**Fig. 5.** Primate Southern blot at moderate stringency washing conditions. Approximately, 5 μg each of male (M) and female (F) genomic DNA preparations from orangutan (*Pongo*), gibbon (*Hylobates*), and rhesus (*Macaca*) were cut with EcoRI, and Southern blotted. The blot was probed with 1F5, washed at moderate stringency conditions (0.5× SSC/65°C), and autoradiographed.

these studies with three different macaques, two gibbons, three orangutans, two gorillas, and two chimpanzees, and have consistently observed the same hybridization pattern (data not shown).

### *Nucleotide Sequence of 1F5*

The deduced sequence of 1F5 is shown in Fig. 6. The 1F5 fragment is 4156 bp long with a GC content of 44.8%. GenBank and EMBL DNA sequence da-

```
YH1F5 4156 bp
   1  GAATTCCTGG GGTCAAGCTA TCTGTTGGTC TCTGCCTCCC AAAGTGTTGG GATTATAGGT
   6  GTGAACCACT GTGCCTAGAC CCCCAAATTA TCTTTTTTTC AGAGACAGTG TGATAGTTCT
 121  CAAACTATCA CCTAACATTC CTGGTAGGAG AGGAAAGATC TCTCTCTTGC CCCAGTCATG
 181  CCTGTCTAAC TACGTGTAAC TAGTTAGCTC GACTTACTGT TATAATTTTT CTCACTGATA
 241  GAATTTTTGC AAAGACACTT TCAATACTAC CAAAATAAAT AGTAATCAAC CACTTGGAAA
 301  GCATTTAGAA GTAACTGAAG TCCCTACTTA CACCAAATGA GGAAAAGCCA AACTACACTG
 361  TGGGACACAC TGTAAGTCCC TGCATCAGTC AGCTGTTGCT GTGCAAACAC CTCCTCAAGA
 421  CTCGGCAGGA TAAAGCAATG ACCATTTATA ATTGCTCACA TGTCTGCAGG TCAGCTGCTT
 481  TTTGGCTGAC CGAGGCAGGG CTTAGCTAGG CTGGCTCTGC AGCGTGTGTC TCTCATTCTC
 541  CTCCTGTGGC AAACAATGGC AGATGACAAT GGCAAATGTG GAAAGCCTTT AAAGAAGAAA
 601  TCTCATAAGG AGTTCCCTGT CACTCCCATT CATCTCAGTC TTTCAGCCAA AGCAAGTGAT
 661  ATGGCCAAAC CCCATATCAA GGGGTGGGGA AATATAATTT GCTTATTTAT TAGGAGCAAA
 721  GTCACATGGT GGTCATTGCA GAATGGGTTG AATGACATAA CCTATCATCA GCCCTGAAGA
 781  AACTCATCAT TCATTTCCTT CCTCAAGACT CCCCTCATCC GGGCATCTCC AGATCTTACT
 841  CTTTCCCACA CTCTGTATCA AACAGTCCAC TCACACTAGC TACCCTGCTT CCTTTGGCAC
 901  CCTCTTTCCA CTAGATAACC TGTTCACTCA GCATTTCTAT TTTCATTTGC ATCTCCACCC
 961  TTTGCCCATT ACCCCAGCAG CACACAGAGG ACCCTGACAG AATAGGCAAA ACCCAGAAGT
1021  TGGGTGGGTA GATAATGTGT GAAAGCTCAC ACGGTTCTGA CGCCTCCTTG ACGCAGCACA
1081  TTACTAGCGA ATGACATGTA TAGGGCCTGG GGTGTTTGCA GAACCAAATC AGGTGTGGAG
1141  GAAATTATAT GTAAATACCT CCATGGGTGT GTGATTGTAT CATACCTCAC CATAACAAAG
1201  CTCAAATGAC ATTTCATTTC CAAATGTGGC TAGGTTTCCC TGTGTTATAT GTCGTTGAAA
1261  GCCAGTATAA ATAAATATCT GTGAAAGTCC TAGCTGACTG CCTGCTTTGT TTCTGCCACT
1321  AAACTGTAAG CTCCTCATGA ACAGAGTGTT TCTGTACAAT TCACTAGCTG TCTCTATCTT
1381  TACCTCACAG CCAATCCACA AATGAACGCT GTCAGCCTTG CCCCCCATCA TATCCCAACC
1441  ATCCATGTCT GGCTACTCCA TTGCTAGTCC AACCGAAGCC CCTTGCATCT CATGCCTGGG
1501  TTATTTCACA GCCTTCTAGG CTGGTAACTC CCGCTGCTAC TGTGCCTGGC TCCAGTCTAT
1561  TCTGTACATG GTGAAGAGGA TCCTTTGAAA AATGGATAAA ATGAGTAGCA AATTGTGTTT
1621  CTCCTGCACT CAGAATCCTC CTCTGGCTCC TCACTGTAAC CTCTGGCTTC AGTAGTTACG
1681  GCAAATATTT AAAAGTCAGC CTGCAGGAAG AACCCTGATT TACTGAGTTT TCCAATGTTA
1741  GTAGTATAAA TACTCATCAT GGCCAATTTT AAGCTGGCTT CTTTCCTTCA CTCCGTCCCT
1801  CCCTCCTCCC CTTCTCCCCT CCTTCTCTTC TCCCCTCTTT CCTTCATTTC CTCCCTCCCT
1861  CCCTCATTTC TTCTGTCCTT CTCTCCTTCC CTTCCTTTCC TTCTCACGTC CTTCTTTCCT
1921  TTCTTCCCTC CTCCCGTCCT TTCTTCCCTA TTCTCTGCCT CCTACTCTAG AATTGACATT
1981  TCATGACAGG CAATCCATGT CTGTCTTAGT TATTGCTTAG AAGTACCTGG CACATAATGG
2041  AAGGCTCAGT ACATATTTCT TGAACAAATA GTTGACTAAT TGAATAGATG ATCCTTTCTT
2101  AATTCTGTAC CAGGATTTAA GAATTATTCA AAGAAGATGT CATTGTTTTA AAATGACCAA
2161  GTACTGCTCA GCACTTTCAA GTTCACAAAG CATGCCTCAT GCATCATCTC TATGATCTGT
2221  ACCCGGGTGC CCATCCTGAG AAGCTGGGAA AGGCTTTTCT CTTCCTCCTC ACTTTACAAA
2281  TGAGGAAACA AAGACTCAGA GGGATGGAGG TTCCCGCTCA AGGCCAAACA GTAAGTATGT
2341  ATCCACAAGT CTGACTCTCA GTCTGGCAGT CTTCTCAGTG GTCTCCAGGA ACTGTTTAGC
2401  TCTTGGGGGG ATCTCATAAG GGAAGCCTGG AGGCTTCAGG GAGAGCTGTC CCAAGGAAGA
2461  GGAGAGAAAC ACATGGTGGC GAGGAAGGAG GATTGACGAC CAGGGAAGGC CTTACGTAGG
2521  CCTCTACCAG GCTGAGCCTC TGGTGTCAGG CCAGTAACAT AGTAATTCAG GTCCCAATCC
2581  TGATTCTGTC TCTAGATTCC CACTAACCTC CCCACAGACA TGGAGCATTG AGCCCCCCTT
2641  CCCCACTTAT AGAGTACCCC AGATAGCATA GGCTCCACAC CTAGCACCCA CCTTCTTGCA
2701  AGTCTATAGC AATGAGCAGC AGAGGCGCAG CTGAGATCCA CAAGCTGAAC CTGGTCCCCA
2761  GGGATAATTT CTTTAGCCCC TGTAGTGTTT AAAAAATTGG GCCAGCATTT ACAAGCAGGA
2821  TATCCCAAGT GGTTGGTTGA TTTTGGTAAA CAGAAAGATC TTGCCACACA TCCTGGCCAT
2881  TCTGGGCTGA CAGCAGTCAT CCACCTGAGA CCAGATAAGT GTTGTCCTGT TCAATGCAGC
2941  CACTCAAACC AGTTCAATGC AGTCCAGTTC AATGCAGTTT CCTATCACCT CTCCAACTTC
3001  CAGGGCAGCC ATTGACTGCA GTTTATTTCC ATGCTGGAAG TTTTACTTGT GTCAAAAAAC
3061  ATACCTCTGA AACCGAAACT TATACCAAAA GAAGAAAAAA TATGTAGGGC TTCATGATTT
3121  TACTACATGG ACTGACAGAA CCCTGAGTTT ACGAGCTTTG GGCTATGCTC AGTTTAAAAT
3181  ACATGCCTGC CTTCCTTCCT TCCTTCCTTC CTTCCTTCCT TCCTTCCTTC CTTCCTTCCT
3241  TCTTTCCTTC CTTCTTCCCT TCTCTTCCTT CCTCCCTCCA TCCCTCCCTT CCTTCCTTCC
3301  TTCCCTCCTT CCTTCCTTCC TTCTTTCCTT CCTTTCTCTC CTTCCTTCCT CCTTCTCTCT
3361  TCTCTTTTTT CTTTTCTTCC CTGAAAGAGT CTCTCTGTTA CCCAGGTATG GTGTATAATA
3421  CATGCTAACA CCACCACACC CAGCTAATTT TTTAATTTCT ATTTTGAAGA AATGAGATCT
3481  TGCCATCTTG CCCAGGCTGG TCTCTAACTC CTGGCCTCAA GTGATCCTCC TGCTTCAGCC
3541  TCCTAAAGCA TTGGGATTGC AGGCATGAGC CACCTAGCCC AGTTAAGTAC ACCTTTTCCT
3601  AGTTCTGTCA ACAATCTCAC AAAGCTGGCA TCCAGAACCC TCCTTTAGGT GAGGATCCTT
3661  GAGTTAAGAG GCGTTGAGGG ACTTGGCCAA GGTTAAATAA CTAGAAGTAG ACACAGAATG
3721  TGAAGATAAC ATCTCATCCA GCGCAGGCAT CCCCATAAAT TGCATTCCCA CCCACACACC
3781  ACCAGTGTGT TCATTTGTAC AGGATGTGTG ACCTGACACA ACAGCCCTTG TGTAATCTGA
3841  CACATTCCCA TTAGGTAGCG ACATAAACCA ACAAAGCTTT CCAGGACAAT GCCTTTGTGG
3901  AAATTGGTTT TCTTGCATTA CTAGGGATTT TAAGGCACAA CAGGTTTTCC CTTCTTTCTG
3961  TCTCTCAAAA AAAAAAAAAA AAAAAAAGTC CTTTCAATAC CCAATATTGC ATGGCTTTCA
4021  TGCCAGGCAT AAAATGAAAT TTGCCATAAT CCATTGCCTA GGGAGAAAGT CGTATGTGTG
4081  TGTCACAACT GACATCTGTT AAATTAAATA AGCAAGCAAT AAGAAGGTCT GAAATATGTT
4141  CTTTGTTCAG GAATTC
```

**Fig. 6.** Nucleotide sequence of 1F5. Only one strand is illustrated although both strands were sequenced. The two polypyrimidine–polypurine segments are underlined and the partial Alu element is double underlined.

tabase searches revealed that the 1F5 sequence has not been previously described. Only limited sequence similarity with other eukaryotic, prokaryotic, and viral sequences was seen. Interestingly, sequence analysis indicated the presence of three regions of potential biological significance in this fragment.

One of these regions contains an Alu-like element (double underlined in Fig. 6), beginning at position 3584 with a right to left orientation. The 1F5 Alu element differs from most other Alu elements described as it appears that only the 5′ (left half) monomer of the typical Alu dimer is present. Figure 7 shows a comparison of the Alu consensus sequence with the opposite strand of 1F5 from position 3584 to 3290. The 1F5 partial Alu shows approximately 75% sequence similarity with the 5′ monomer of the Alu consensus sequence (Deininger et al. 1981).

Significantly, unlike most Alu elements, the 1F5 partial Alu is not flanked by direct repeats.

Besides the partial Alu element, 1F5 has two other sequences with rather interesting arrangements. Both of these regions show striking strand asymmetry in that one strand is composed primarily of pyrimidine residues (CT) and the opposite primarily of purine residues (GA). Such regions have been termed polypyrimidine–polypurine (PUPPY) segments (Birnboim et al. 1979). Both of these segments are approximately 190 bp long. The first such segment is located at position 1778–1968 (underlined in Fig. 6). The strand shown is composed of 95% pyrimidine residues. No obvious repetitive pattern is seen in the first PUPPY segment of 1F5. The second PUPPY segment is located at position 3190–3382 (underlined in Fig. 6). In this case, 99.5% of the residues are pyrimidines. The second segment does show a pattern of repetition in that the sequence CCTT is repeated 13 times at the beginning of the sequence element (Fig. 6). Beyond this region, however, the second element does not show an obvious repetitive pattern.

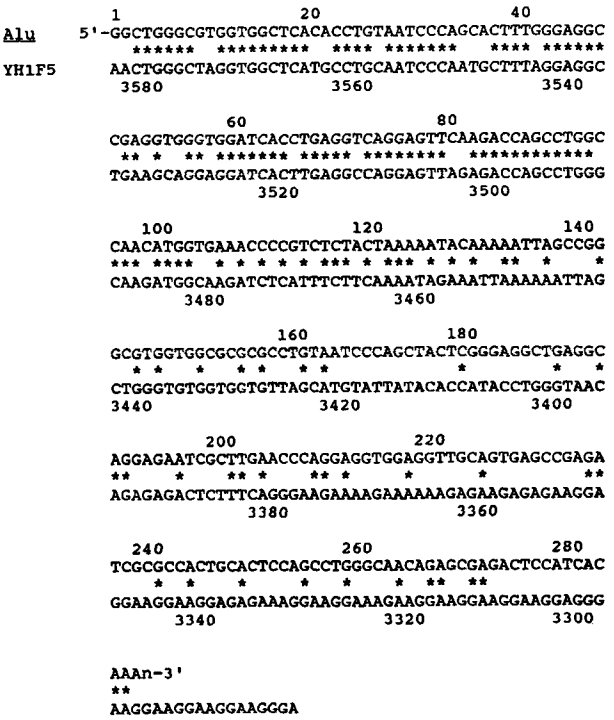


**Fig. 7.** Comparison of the Alu consensus sequence with the 1F5 partial Alu. The Alu consensus is the upper strand and the 1F5 partial Alu sequence is the lower. Conserved bases are depicted by *. The numbers indicate positions in the Alu consensus or 1F5 sequences.

## Discussion

### *Evolutionary Conservation of 1F5*

A number of studies have revealed conservation of Y-unique restriction fragments in human, chimpanzee, and gorilla, whereas some divergence of these fragments was noted in orangutan (Page et al. 1984; Koenig et al. 1985; Erickson 1987; Yen et al. 1988). However, in the lesser apes (gibbons) and macaques, a greater degree of divergence was seen. Interestingly, the human Y-unique sequences are frequently found on the X chromosome or autosomes of other primates (Page et al. 1984; Koenig et al. 1985; Erickson 1987; Yen at al. 1988).

The data presented here suggest that the male-specific 1F5 fragment has an ancient association with the Y chromosome. A fragment with sequence similarity to 1F5 is conserved in chimpanzee, gorilla, orangutan, and macaque genome. An exception to the conservation of the 1F5 sequence in primates is seen in gibbon, which possesses only the 6.7-kb male–female-shared band. No male-specific hybridizing band(s) were seen (Fig. 4). The absence of a 4.1-kb male-specific band in the gibbon could be explained by the loss of 1F5 homologous sequences by a deletional event. This event probably occurred after the divergence of the gibbon line from the common line to the great apes and human some 18–22 million years ago (Andrews 1986). However, alternative explanations are also plausible.

Hybridization data presented here suggest >80% sequence similarity between 1F5 and the 6.7-kb fragment detectable under moderately stringent washing conditions (Fig. 1). The 6.7-kb fragment is apparently localized to autosomes as judged by its absence in the hamster/human hybrids containing either the human X or Y chromosome (Fig. 2). This fragment is conserved in all the primates examined. It is possible that the origin of 1F5 may have involved transposition of the 6.7-kb fragment followed by its rearrangement on the primate Y chromosome.

### *Alu Family of Repetitive Elements*

The major repetitive DNA family of the human genome consists of Alu elements (Deininger et al. 1981). The 300,000–500,000 members of the Alu family consist of short interspersed repetitive DNA elements (SINE family), which are widely distributed in the human genome (Jelinek and Schmid 1982). The human Alu sequence is dimeric, composed of two monomer units of similar, but not identical, sequences arranged in tandem. The sequence similarity between an individual Alu family member to the Alu consensus sequence is variable, but typically averages 87%. The length of the entire repeat is approximately 300 bp (Deininger et al. 1981). The sequence at the 3′ end is demarcated by a region rich in adenosine residues, and most Alu

family members are flanked by short direct repeats (Schmid and Jelinek 1982).

It has been proposed that the Alu sequences on the human Y chromosome have significantly diverged relative to average genomic Alu sequences (Smith et al. 1987). Based on their sequence data, Smith et al. (1987) further proposed that the Alu element in 2.4-kb Y repeats is similar to the type II Alu repeats described in *Galago* (Daniels and Deininger 1983). Additionally, less well-conserved Alu elements have been described by Gibbs et al. (1987) and Ellis et al. (1989). The 1F5 Alu-like element shows partial sequence similarity only with the left-half monomer and, thus, is clearly distinct from the Alu sequences described earlier (Gibbs et al. 1987; Smith et al. 1987; Ellis et al. 1989). Evidence from the sequence of 2D6 (another human Y-chromosomal fragment; Rasheed et al. 1991) also shows that the presence of truncated and divergent Alu-like sequences may not be uncommon in this chromosome.

### *Polypyrimidine–Polypurine (PUPPY) Sequences*

In addition to the partial Alu, 1F5 also contains two 190-bp regions of PUPPY sequences displaying obvious strand asymmetry (Fig. 6). Like Alu sequences, PUPPY sequences are apparently widely distributed in the DNA of higher organisms (Birnboim et al. 1979). These sequences have been postulated to function in transcriptional regulation as they have been shown to take on unusual non-B DNA structures such as intramolecular triplexes, which make them susceptible to attack by single strand-specific endonucleases (Wells et al. 1988). Additionally, these types of sequences have been implicated in recombinational events (Hoffman-Liebermann et al. 1986; Konopka 1988).

A role for the PUPPY sequence in 1F5 is unclear as no transcriptional elements were identified. Alternatively, it is possible that these sequences may be recombinational sites. In this context, it may be important to note that one of the two PUPPY sequences (99.5% pyrimidines) is only 60 bp from the 3′ end of the 1F5 Alu left-half monomer (Fig. 7). Therefore, any possible recombinational event, which might have led to the formation of the partial Alu element, may have involved the polypyrimidine–polypurine segment.

*Acknowledgments.* The authors gratefully acknowledge constant encouragement and support from Dr. W. Wertelecki during the course of these studies. The computer analysis of the DNA sequence and the database (GenBank and EMBL) searches were aided by Dr. David Wood, Dr. Gary Daniels, Mr. Carlos Monteiro, and the Computer Support group of the University of South Alabama. We also thank Dr. Rajani Joshi, Sheela Bhatnagar, Sarla Sharma, C.M. Clayton, and B. Allen for their help during the course of these studies. The secretarial assistance of Mrs. Sheila White is gratefully acknowledged. H.O. is supported by a Basil O'Connor Starter Grant (5-585) from the March of Dimes/Birth Defects Foundation, and from the R.C. Phillips Research and Education Unit, Children's Medical Services, Department of Health and Rehabilitative Services, Florida. Y.M.B. is a recipient of the Research Career Development Award from NICHHD, NIH.

## References

Affara NA, Florentin L, Morrison N, Kwok K, Mitchell M, Cook A, Jamieson D, Glasgow L, Meredith L, Boyd E, Ferguson-Smith MA (1986) Regional assignment of Y-linked DNA probes by deletion mapping and their homology with X-chromosome and autosomal sequences. Nucleic Acid Res 14:5253–5373.

Andrews P (1986) Fossil evidence on human origins and dispersal. Cold Spring Harbor Symp Quant Biol 51:419–428.

Birnboim, HC, Sederoff RR, Paterson MC (1979) Distribution of polypyrimidine–polypurine segments in DNA from diverse organisms. Eur J Biochem 98:301–307

Bishop CE, Guellaen G, Geldwerth D, Fellous M, Weissenbach J (1984) Extensive sequence homologies between Y and other human chromosomes. J Mol Biol 173:403–417

Bishop CE, Weissenbach J, Casanova M, Bernheim A, Fellous M (1985) DNA sequences and analysis of the human Y chromosome. In: Sandberg A (ed) The Y chromosome, part A: basic characteristics of the Y chromosome. Alan R Liss, New York, pp 141–176

Bobrow M, Pearson P, Pike M, El-Alfi O (1971) Length variation in the quinacrine-binding segment of human Y chromosomes of different sizes. Cytogenetics 10:190–198

Buhler EM (1985) Clinical and cytologic impact of Y chromosome abnormalities. In: Sandberg A (ed) The Y chromosome, part B: clinical aspects of Y chromosome abnormalities. Alan R Liss, New York, pp 61–93

Cooke HJ, McKay RDG (1978) Evolution of a human Y chromosome specific repeated sequence. Cell 13:453–460

Daniels GR, Deininger PL (1983) A second major class of *Alu* family repeated DNA sequences in a primate genome. Nucleic Acids Res 11:7595–7610

Davis LG, Dibner MD, Battey JF (1986) Basic methods in molecular biology. Elsevier, New York

Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol 151:17–33

Ellis NA, Goodfellow PJ, Pym B, Smith M, Palmer M, Frischauf A-M, Goodfellow PN (1989) The pseudoautosomal boundary in man is defined by an *Alu* repeat sequence inserted on the Y chromosome. Nature 337:81–84

Erickson RP (1987) Evolution of four human Y chromosomal unique sequences. J Mol Evol 25:300–307

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Gibbs PE, Zielinski R, Boyd C, Dugaiczyk A (1987) Structure, polymorphism, and novel repeated DNA elements revealed by a complete sequence of the human $\alpha$-fetoprotein. Biochemistry 26:1332–1343

Harper ME, Saunders G (1981) Localization of single copy DNA sequences on G-banded human chromosomes by in situ hybridization. Chromosoma 83:431–439

Harper ME, Ullrich A, Saunders GF (1981) Localization of the human insulin gene to the distal end of the short arm of chromosome 11. Proc Natl Acad Sci USA 78:4458–4460

Henikoff S (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol 155:156–165

Hoffman-Liebermann B, Liebermann D, Troutt A, Kedes LH, Cohen SN (1986) Human homologs of TU transposon sequences: polypurine/polypyrimidine sequence elements that can alter DNA conformation in vitro and in vivo. Mol Cell Biol 6:3632–3642

Jelinek WR, Schmid CR (1982) Repetitive sequences in eukaryotic DNA and their expression. Annu Rev Biochem 51: 813–844

Koenig M, Moisan JP, Heilig R, Mandel J-L (1985) Homologies between X and Y chromosomes detected by DNA probes: localisation and evolution. Nucleic Acids Res 13:5485–5501

Konopka AK (1988) Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells. Nucleic Acids Res 16:1739–1758

Kunkel LM, Smith KD, Boyer SH (1979) The organization and heterogeneity of sequences within a repeating unit of human Y chromosome DNA. Biochemistry 18:3343–3352

Lai EC, Woo SLC, Dugaiczyk A, O'Malley BW (1979) The ovalbumin gene: alleles created by mutations in the intervening sequences of the natural gene. Cell 16:201–211

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor NY

Page D, Harper ME, 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 311:119–123

Palmer MS, Sinclair AH, Berta P, Ellis NA, Goodfellow PN, Abbas NE, Fellous M (1989) Genetic evidence that ZFY is not the testis-determining factor. Nature 342:937–939

Pustell J, Kafatos FC (1984) A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res 12:643–655

Rasheed BKA, Whisenant EC, Fernandez R, Ostrer HO, Bhatnagar YM (1991) A Y-chromosomal DNA fragment is conserved in human and chimpanzee. Mol Biol Evol (in press)

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5468

Schmid CR, Jelinek WR (1982) The *Alu* family of dispersed repetitive sequences. Science 216:1065–1070

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf A, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240–244

Smith KD, Young KE, Talbot CC, Schmeckpeper BJ (1987) Repeated DNA of the human Y chromosome. Development 101[Suppl]:77–92

Stewart AD (1983) The role of the Y-chromosome in mammalian sexual differentiation. In: Johnson MH (ed) Development in mammals, vol. 5. Elsevier, Amsterdam, pp 321–367

Vergnaud G, Page D, Simmler M-C, Brown L, Rouyer F, Noel B, Botstein D, de la Chapelle A, Weissenbach J (1986) A deletion map of the human Y chromosome based on DNA hybridization. Am J Hum Genet 38:109–124

Weissenbach J, Goodfellow PN, Smith KD (1989) Report of the committee on the genetic constitution of the Y chromosome. Cytogenet Cell Genet 51:438–449

Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F (1988) The chemistry and biology of unusual DNA structures adopted by oligopurine–oligopyrimidine sequences. FASEB J 2:2939–2949

Willard HF (1985) Molecular organization of repeated DNA sequences on the human Y chromosome. In: Sandberg A (ed) The Y chromosome, part A: basic characteristics of the Y chromosome. Alan R Liss, New York, pp 125–140

Yen PH, Marsh B, Allen E, Tsai SP, Ellison J, Connolly L, Neiswanger K, Shapiro LJ (1988) The human X-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution. Cell 55:1123–1135

Yunis JJ (1976) High resolution of human chromosomes. Science 191:1268–1271

Received November 16, 1990/Revised February 7, 1991