Am. J. Hum. Genet. 46:950–955, 1990

Strong Linkage Disequilibrium Between the *XY274* Polymorphism and the Pseudoautosomal Boundary

Nathan Ellis,* Judy Kidd,† Paul J. Goodfellow,*,[1] Kenneth Kidd,† and Peter N. Goodfellow*

*Imperial Cancer Research Fund, Lincoln's Inn Fields, London, England; and †Department of Human Genetics, Yale University School of Medicine, New Haven, CT

Summary

The pseudoautosomal boundary is defined by an *Alu* repeat element on the Y chromosome. The *Alu* element is found on all Y chromosomes and on no X chromosomes, establishing it as part of Y-specific sequences. Distal to the *Alu* element, sequences from the X and Y are strictly homologous, suggesting that the boundary is formed by an abrupt break in sequence homology. Further investigation of the function of the boundary has been undertaken by examining the population structure of an *Msp*I restriction-site polymorphism (*XY274*), which is located 274 bp distal to the *Alu* insertion site. Southern blot and polymerase chain reaction analysis demonstrate fixation of the *high* allele (noncutting or AT base pair) of *XY274* on the Y chromosome in most populations, while a full range of *high* allele frequencies is found on the X chromosomes of different populations. Two exceptions to fixation on the Y chromosome were found in African populations. The level of linkage disequilibrium suggests that the first few hundred base pairs of the pseudoautosomal region on the Y chromosome share a single common origin more recent than the origin of the species.

Introduction

The mammalian Y chromosome is composed of two genetically distinct parts: one functions in sex determination and male fertility, the other is presumed to function in the segregation of the Y chromosome from the X chromosome during male meiosis. The pseudoautosomal region is a segment of strict homology between the sex chromosomes, where they pair and exchange material by homologous recombination (Weissenbach et al. 1987). In humans, it is located on the distal tips of the short arms of the X and Y chromosomes. Long-range restriction mapping has been used to estimate the size of the pseudoautosomal region at 2.6 million bp (Brown 1988; Petit et al. 1988). Genetic and physical distance have been compared at four evenly spaced loci in the pseudoautosomal region. Each segment maintains an average pseudoautosomal rate of recombination such that 1% is equivalent to 5×10^4 bp (Petit et al. 1988)—a rate 20-fold greater than the genome-wide average, where 1% is equivalent to 10^6 bp.

Recombination in the pseudoautosomal region is limited at its proximal end by the pseudoautosomal boundary. The boundary is the interface between pseudoautosomal and sex chromosome–specific regions. It functions to keep the two segments distinct and to prevent recombination entering into sex chromosome–specific DNA. The boundary in man was isolated in a chromosome walk initiated at flanking loci: the pseudoautosomal gene *MIC2* and the Y-specific locus *DYS104* (Ellis et al. 1989). The boundary region was identified by restriction mapping as a transition between pseudoautosomal and sex chromosome–specific sequences. Sequencing around the transition point identified an abrupt break in sequence homology (fig. 1). X-Y homologous sequences showing high (>99%) homology extended up to a single base pair. Only two differences in 245 bp were found between X and Y sequences. At the end of the homology between the X and Y chromosomes, an *Alu* repeat element inserted

Received October 23, 1989; revision received January 22, 1990.

Address for correspondence and reprints: Dr. Nathan Ellis, Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, England.

1. Present address: Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

0002-9297/90/4605-0013$02.00

on the Y chromosome was found. Proximal to the *Alu* insertion, partial (77%) homology between the X and Y chromosomes extended for 225 bp, after which homology was lost. We hypothesized that about 25 million years ago the *Alu* insertion moved the boundary from its former position 225 bp proximal to the *Alu* insertion sites to its present position at the insertion site. This hypothesis explains the strict homology distal to *Alu,* but only partial homology proximal to *Alu.* The change in the region of partial homology acts as a clock for the time of *Alu* insertion.

Southern blot analysis determined that the *Alu* is always present on the Y chromosome and never on the X, implying that the *Alu* segregates with Y-specific sequences. However, what defines the boundary formally is whether recombination occurs up to the insertion site. One way to approach this question is to examine polymorphisms close to the boundary region and compare haplotypes on the Y chromosome to those on the X chromosome. We have found a two-allele polymorphism (called *XY274*) which is 274 bp distal to the insertion site. The polymorphism has been tested in Caucasians, Oceanic populations, australoids, Amerindians, and Africans. Genotype frequencies indicate strong linkage disequilibrium between *XY274* and the insertion site. Separation of the Y-associated allele from the Y chromosome was observed in two African populations. The data indicate that the first few hundred base pairs distal to *Alu* on the Y chromosome share a common origin more recent than the origin of *Homo sapiens*.

Material and Methods

DNA Samples

DNA samples from Caucasians were gifts from Julia Bodmer and Nigel Spurr. Samples from Western Highland Papua New Guinea and Tubuai Island (Polynesia) were gifts from J. Martinsen (Flint et al. 1989). Samples from Kalahari !Kung San (formerly "Bushmen"; the !Kung constitute the largest group of the northern division of the San. The ! symbol is an orthographic convention representing one of a number of click sounds found in Khoisan ["Bush" and "Hottentot"] languages) and Southern African Bantu-speaking blacks were gifts from T. Jenkins (Ramsay and Jenkins 1988). Samples from Pygmy have been described in Bowcock et al. (1987). Rondonia Surui and Karitiana are Amazon-basin Indians collected by Dr. F. Black; Campeche Mayan samples from the Yucatán in Mexico were collected by Dr. K. Weiss.

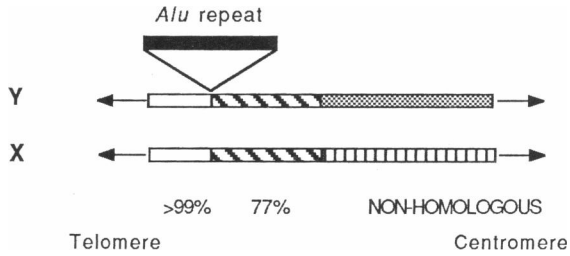

Figure 1 Representation of the model of the pseudoautosomal boundary. An *Alu* repeat element defines the boundary on the Y chromosome (*PABY*). The same position on the X chromosome is the X boundary (*PABX*). Pseudoautosomal sequences lie distally, and sex-specific sequences lie proximally. A small region of partial (77%) homology proximal to the *Alu* repeat is a relic of the insertion event which created the present-day boundary about 25 million years ago.

Southern Blot Analysis

DNA samples were digested with *Msp*I, separated by electrophoresis through 0.7% agarose gels (SeakemME), and transferred to GeneScreen Plus™ (New England Nuclear) or Hybond N Plus (Amersham) according to a method described by Goodfellow et al. (1986). The probes Hf0.2 and RsY0.55 have been described by Ellis et al. (1989). Fragments were isolated on NA45 paper and labeled with ^{32}P by the multiprime method (Feinberg and Vogelstein 1983). Hybridization was in 1 M NaCl, 10% dextran sulfate, 1% SDS, 0.25 mg sheared and denatured salmon testes DNA/ml with 2–10 ng labeled probe/ml. Filters were washed at 0.2 × SSC, 0.2% SDS at 65°C for up to 2 h. Autoradiography was to Kodak XAR-5 film with Lightning Plus intensifier screens.

Polymerase Chain Reaction (PCR) and Analysis

PCR was used to amplify X and Y boundaries and *XY274*. Three primers were synthesized: one pseudoautosomal (CTGAGAGTGGAAGTGTCGCAG), one Y specific (GTACTACCTTTAGAAAACTAGTATTTTCCC) and one X specific (CTGCAGAAACAAGCTCATCAGCGTGACTAT). The 30-mer oligos were chosen at first because Y-specific sequences are very AT rich. Reactions used 1 μg of genomic DNA in 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris HCl pH 8.0, 0.1% gelatin, 200 μM each of dATP, dGTP, dCTP, and dTTP and 1.25 units *Taq* polymerase or AmpliTaq (Cetus) in a 50-μl volume overlaid with 1–2 drops of mineral oil (Sigma). The reaction was incubated at 94°C for 1 min, 54°C for 1 min, and 72°C for 1 min for 30 cycles. A final incubation at 72°C for 6 min was added to the end of the last cycle. A 10-μl sample of each PCR product was subsequently digested with 3 units of *Msp*I at

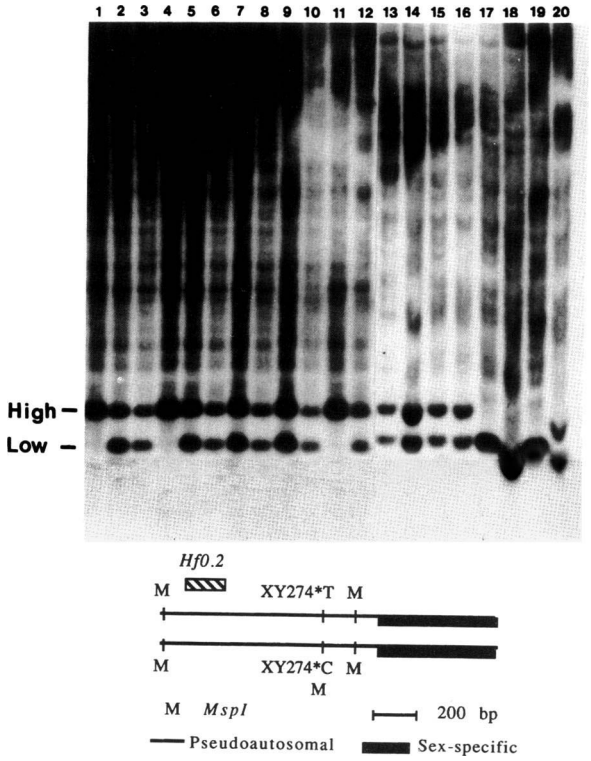

Figure 2 Polymorphism in X-Y homologous sequences, defined 274 bp distal to the *Alu* insertion site (*XY274*). Probe Hf0.2 distinguishes 1.0-kb and 0.8-kb fragments in Southern blots of *Msp*I-cut DNA. *XY274* was sequenced from the Y chromosome (pSACY3.3) and the X chromosome (pSACX4.5), and it arose from a C-to-T transition in the *Msp*I recognition site (CCGG/CTGG). Lanes 1–12, Tubuai males; lanes 13 and 14, Rondonia Surui males; lanes 15, 16, and 20, Pygmy males; lanes 17–19, Pygmy females.

37°C for 1 h and was analyzed by electrophoresis through 1% agarose gels.

Results

The Y boundary region (*PABY*) is carried on the plasmid pSACY3.3, and the X boundary region (*PABX*) is carried on pSACX4.5. Sequence analysis defined two base differences between the X and Y chromosomes in the 245 bp distal to the *Alu* insertion site. A third difference was found by restriction-enzyme analysis of pSACY3.3 and pSACX4.5 cut with *Msp*I. The difference mapped in X-Y homologous sequences within 300 bp of the insertion site. Southern blot analysis with a probe derived from X-Y homologous sequences, i.e., probe Hf0.2, reveals the difference as an RFLP in genomic DNA samples cut with *Msp*I (fig. 2). Two fragments are seen: one 1.0 kb, the other 0.8 kb. The RFLP is defined at the proximal *Msp*I site because hybridization of *Msp*I-cut samples with a probe defining sequences on the next distal *Msp*I fragment, HIII0.55, reveals nonpolymorphic DNA fragments (data not shown). Subsequent sequence analysis of pSACY3.3 and pSACX4.5 showed the polymorphism to be located 274 bp distal to the boundary and to arise from a C-to-T transition of the second base of the *Msp*I recognition site (CCGG/CTGG). We have called the polymorphic locus "*XY274*." The 1.0-kb *Msp*I fragment represents the *high* allele (*XY274*T*) and is determined by a T at *XY274;* the 0.8-kb *Msp*I fragment represents the *low* allele (*XY274*C*) and is determined by a C at *XY274*.

In order to study the population structure of the *XY274* polymorphism, the allele and genotype frequencies in males and females were measured in northern Europeans (Caucasians), Western Highland Papuans (australoids), Tubuai Polynesians (Oceanics), Rondonia Surui, Mayans and Karitiana (Amerindians), Zaire and Central African Republic Pygmies, San, and Southern African Bantu-speaking peoples (Africans). Table 1 shows the genotype frequencies in males and females and the frequency of the *high* allele on the X (P_x) and the Y (P_y) chromosome as determined by maximum-likelihood estimation.

Each of the major population groups has a different X-chromosome frequency, P_x, ranging from 0% to 100%. Papuans have nearly fixed the *high* allele on the X (P_x = .92), whereas in the three Amerindian populations the *high* allele was either not found, or found infrequently, on the X chromosome. Northern Europeans have the intermediate value of .54, and Tubuai Islanders are between northern Europeans and Amerindians, with a value of .18. African blacks show a range of P_x values. Pygmies and Bantu-speaking Africans have P_x values of .11 each, while the San have a P_x value of .32.

P_y values are strikingly uniform. In all but two populations, the *high* allele appears fixed on the Y chromosome. This is seen most obviously in the Rondonia Surui and Karitiana, where all the females are homozygous *low* and all the males are heterozygous. The important class to note in males is the number of homozygous *low* individuals. This is zero in all populations except in San, where we find four *low* individuals, and in Pygmies, where we find six. The San population has a P_y value of .72, and the Pygmy population a value of .61. Thus, by this RFLP analysis, we have not been able to demonstrate the presence of the *low* allele on the Y chromosome, except in San and Pygmy.

Table 1

Genotype and Allele Frequencies of the *XY274* Polymorphism

Population and Genotype	No. of Males	No. of Females	P_x[a]	P_y[b]
Northern European:				
High/high	38	17		
High/low	30	20	.54 (170)	1.00 (68)
Low/low	0	14		
Western Highland Papuan:				
High/high	12	5		
High/low	1	1	.92 (25)	1.00 (13)
Low/low	0	0		
Tubuai Polynesian[c]:				
High/high	3	. . .		
High/low	14	. . .	.18 (17)	1.00 (17)
Low/low	0	. . .		
Rondonia Surui:				
High/high	0	0		
High/low	19	0	.0 (49)	1.00 (19)
Low/low	0	15		
Mayan:				
High/high	1	0		
High/low	8	2	.048 (63)	1.00 (9)
Low/low	0	25		
Karitiana:				
High/high	0	0		
High/low	22	0	.0 (82)	1.00 (22)
Low/low	0	30		
Pygmy:				
High/high	2	1		
High/low	9	6	.11 (95)	.61 (17)
Low/low	6	32		
Kalahari San:				
High/high	7	1		
High/low	8	5	.32 (47)	.72 (19)
Low/low	4	8		
Southern African Bantu:				
High/high	4	0		
High/low	22	3	.11 (64)	1.00 (26)
Low/low	0	16		

[a] Frequency of *high* on the X chromosome, as calculated by maximum-likelihood methods. Numbers in parentheses are the number of X chromosomes tested.

[b] Frequency of *high* on the Y chromosome, as calculated by maximum-likelihood methods. Numbers in parentheses are the number of Y chromosomes tested.

[c] Since only males were tested, there are two equally likely solutions for P_x and P_y that are symmetric. P_x = .18 or 1.0, and, similarly, P_y = 1.0 or .18. Given the run of the data, we have chosen the solution where P_y = 1.0.

In order to confirm both the complete allele association between *high* at *XY274* and the Y boundary in northern Europeans and the lack of it in San, we determined haplotypes of *XY274* and the pseudoautosomal boundary by PCR. A 21-mer derived from X-Y homologous sequences 156 bp distal to *XY274* was used in conjunction with either a Y-derived 30-mer 702 bp proximal to the insertion site (Y-oligos) or an X-derived 30-mer 526 bp proximal to the insertion site (X-oligos) (fig. 3). In males, PCR with Y-oligos amplifies a 1.1-kb Y-specific fragment and PCR with X-oligos amplifies a 0.95-kb X-specific fragment. When the product is digested with *Msp*I, the *high* allele is recognized by a 336-bp fragment and the *low* allele by 156-bp and 180-

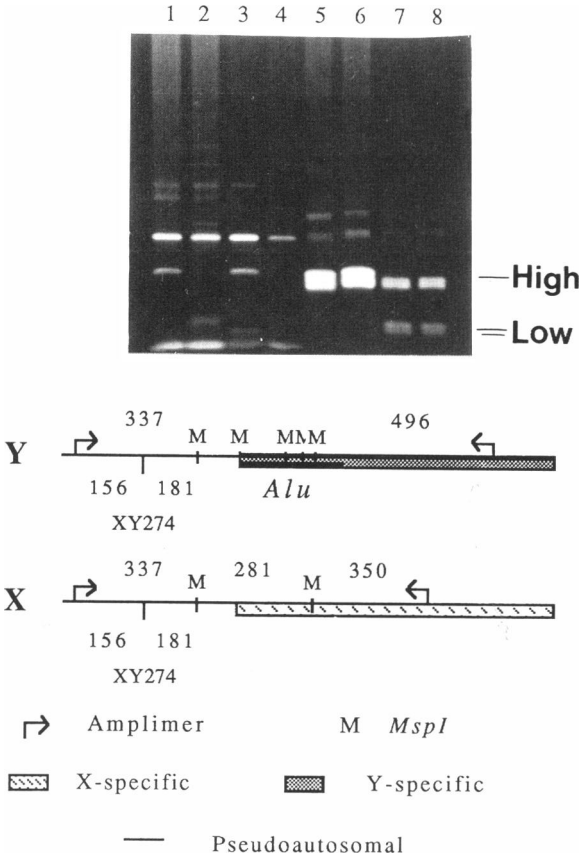

Figure 3 Identification of *XY274* polymorphism by PCR. PCR using oligos on both sides of the insertion site amplifies sex-specific fragments. When the products are digested with *Msp*I and electrophoresed on 1% agarose gels, the *XY274* polymorphism is scored either *high* by the presence of a 337-bp fragment or *low* by the presence of 156-bp and 181-bp fragments. Amplified DNAs are from San males. Lanes 1–4, *Msp*I-digested products from Y-oligo reactions; lanes 5–8, *Msp*I-digested products from X-oligo reactions. Lanes 1 and 5, JR046; lanes 2 and 6, JR033; lanes 3 and 7, JR029; lanes 4 and 8, JR013.

bp fragments usually indistinguishable at the resolution of our gels. In Y-oligo reactions, Y-specific fragments of 496 bp and several indistinguishable fragments smaller than 150 bp are seen. In X-oligo reactions, comigrating X-specific fragments of 350 bp and 281 bp are seen. PCR analysis of females with X-oligos allows identification of all three genotypes.

Table 2 shows the number of each haplotype in northern Europeans and in San and calculates P_x and P_y as determined by PCR analysis. The *high* allele is completely associated with the Y boundary in northern Europeans (P_y = 1.0). In San the *high* allele associates with *PABY* in a little over half the Y chromosomes (P_y = .57). The P_x and P_y values are not significantly different from those determined by maximum-likelihood estimates.

Table 2

PCR Analysis of *XY274* and Boundary Haplotypes

Haplotype	Northern European	San
(*High, PABY*)	19	8
(*Low, PABY*)	0	6
(*High, PABX*)	13	14
(*Low, PABC*)	9	26
P_y	1.0	.57
P_x	.59	.35

Finally, PCR analysis directly demonstrates the structures of *PABY* and *PABX* by the size of uncut and *Msp*I-digested X-oligo and Y-oligo products. The results extend the number of individuals assayed in the boundary region. With the additional 33 Y-oligo and 62 X-oligo products, over 160 *PABX*s and 100 *PABY*s have been examined, and in all cases the *Alu* repeat element is found on the Y chromosome and not on the X chromosome.

Discussion

We have surveyed polymorphism at the locus *XY274* in nine populations from the following distantly related groups: Caucasians, australoids, Oceanic populations, Amerindians, and Africans. The northern Europeans and Southern African Bantu-speaking blacks are both highly urbanized populations and therefore may be expected to be heterogeneous. The other populations are small, isolated groups which probably represent samplings of larger groups.

The P_x value is a characteristic of each of the major population groups. P_x values range from close to 1.0 in Papuans to 0 in Amerindians. Fixation of the *low* allele on the X is found in Rondonia Surui and Karitiana (P_x = 0), but in Mayans P_x = .048, which is consistent with the approximately 10% admixture with northern Europeans that is found in this population (Dr. K. Weiss, personal communication). In contrast, the *high* allele is fixed on the Y chromosome, resulting in strong linkage disequilibrium between *XY274* and the pseudoautosomal boundary. The greatest number of chromosomes have been collected from the northern European population. In females the frequency of *high* and *low* alleles is close to .5, and the three genotypes occur at Hardy-Weinberg frequencies. However,

in males there are only two classes in equal numbers: homozygous *high* and heterozygotes. Since half of the X chromosomes carry *high* and half carry *low,* it is inferred that the Y chromosome carries the *high* allele at or near fixation.

Two populations proved to be exceptional. The Kalahari San have a P_y equal to .72, and Pygmies have a frequency of P_y = .61. Restriction-enzyme analysis of X- and Y-specific boundary fragments amplified by PCR demonstrated complete association of the *high* allele and the Y boundary in northern Europeans and confirmed the presence of the *low* allele on the Y chromosomes in San. No product of a crossover has been identified among non-African Y chromosomes.

Linkage disequilibrium can arise in many ways—e.g., by admixture, genetic drift, selection, or recombination suppression. Genetic drift arises from small effective population sizes, so that the subsequent generations become a sampling of a small number of chromosomes. Since the Y chromosome is inherited through the male sex alone, drift specifically affecting the Y chromosome could occur if the population were founded by a small number of males. The total amount of genetic variation need not be severely reduced if the number of females is sufficiently large. Recombination suppression, on the other hand, lengthens the amount of time required to separate the *high* allele from *PABY.* It is likely the *XY274* polymorphism predates diversification of modern *Homo sapiens* because it is present in all the population groups. Current estimation of the first major divergence (African/non-African split) places the date about 90,000 years ago (Cavalli-Sforza et al. 1988). Given both the small amount of time San and Pygmies have been separated from all other groups and a rate of mutation at a CpG site estimated at 5.4×10^{-9}/year (Cooper and Krawczak 1989), it is unlikely that either *high* or *low* alleles on San and Pygmy Y chromosomes arose by new mutation. Therefore, it may be surmised that the (*low, PABY*) haplotype was present in the ancestral population and subsequently was lost from the other branches of the species by genetic drift. Whether there was strong disequilibrium between *XY274* and the boundary in the ancestral population is not known, nor is it necessary to hypothesize recombination suppression to explain absence or presence of (*low, PABY*) chromosomes in present-day populations. The data suggest that a single class of Y chromosome migrated out of Africa about 90,000 years ago and that present-day non-African males have inherited it.

In summary, we have examined the population structure of *XY274* in nine populations and found complete linkage disequilibrium between *XY274* and the pseudoautosomal boundary in all but two populations. Six individuals in San (table 2), and at least six Pygmy individuals (table 1), have a (*low, PABY*) haplotype (table 2). It is concluded that the first few hundred base pairs of the Y chromosome distal to the *Alu* insertion site in some African and in all non-African populations share a common ancestor originating at the time of the African/non-African split.

Acknowledgments

We would like to thank Julia Bodmer, Nigel Spurr, Jeremy Martinsen, and Trefor Jenkins for valuable discussion and invaluable donations of human DNA samples. We thank Kim Richardson for typing the manuscript.

References

Bowcock AM, Bucci C, Herbert JM, Kidd JR, Kidd KK, Friedlaender JS, Cavalli-Sforza LL (1987) Study of 47 DNA markers in five populations from four continents. Gene Geogr 1:47–64

Brown WRA (1988) A physical map of the human pseudoautosomal region. EMBO J 7:2377–2385

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci USA 85:6002–6006

Cooper DN, Krawczak M (1989) Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Hum Genet 83:181–188

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

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

Flint J, Boyce AJ, Martinson JJ, Clegg JB (1989) Population bottlenecks in Polynesia revealed by minisatellites. Hum Genet 83:257–263

Goodfellow PJ, Darling SM, Thomas NS, Goodfellow PN (1986) A pseudoautosomal gene in man. Science 234: 740–743

Petit C, Levilliers J, Weissenbach J (1988) Physical mapping of the human pseudo-autosomal region: comparison with genetic linkage map. EMBO J 7:2369–2376

Ramsay M, Jenkins T (1988) Alpha-globin gene clusters haplotypes in the Kalahari San and Southern African Bantu-speaking blacks. Am J Hum Genet 43:527–533

Weissenbach J, Levilliers J, Petit C, Rouyer F, Simmler M-C (1987) Normal and abnormal interchanges between the human X and Y chromosomes. Development 101 [Suppl]: 67–74