Gene, 165 (1995) 191–198

GENE 09217

PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis

(DNA polymorphism; repeated sequence; individual profiling; human diversity; male lineage; human evolution)

Fabrício Rodrigues Santos[a], Sergio D.J. Pena[a] and Chris Tyler-Smith[b]

[a]*Universidade Federal de Minas Gerais, Departamento de Bioquimica, Belo Horizonte MG, Brazil; and* [b]*CRC Chromosome Molecular Biology Group, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK*

Received by G. Bernardi: 1 April 1995; Revised/Accepted: 3 July/12 July 1995; Received at publishers: 20 July 1995

SUMMARY

We have developed a system for revealing informative and useful haplotypes for the human Y chromosome using PCR. Variant alphoid satellite DNA subunits were amplified and analysed by digestion with *Hin*dIII to score a restriction site polymorphism, or on polyacrylamide gels to reveal 13 heteroduplex haplotypes. Heteroduplexes are double-stranded DNA molecules containing mismatches; the haplotype is the combination of alleles on the same chromosome. Structural studies showed that the heteroduplexes analysed here were formed from loci at the left (short arm) and right (long arm) edges of the centromeric alphoid array which differed by a 4-bp insertion/deletion and several point mutations. Consequently, many haplotypes may have arisen only once and are useful for evolutionary studies.

INTRODUCTION

Y-chromosome DNA polymorphisms are potentially useful in two areas of human biology. In evolutionary studies they allow paternal lineages to be identified (Spurdle and Jenkins, 1992a) and in individual profiling their male specificity and high exclusion power may make them the markers of choice for some kinds of forensic or paternity testing (Santos et al., 1993a). However, their potential has been largely unrealised because few suitable polymorphic markers have been found despite systematic searches (Jakubiczka et al., 1989; Malaspina et al., 1990; Spurdle and Jenkins, 1992b). Early reports described conventional RFLPs which showed only limited variability (Casanova et al., 1985) or complex patterns which are poorly understood (Lucotte and Ngo, 1985). Pulsed-field gel electrophoresis (PFGE) has subsequently revealed a large amount of variation (Oakey and Tyler-Smith, 1990; Jobling, 1994; Mathias et al., 1994) but PFGE is not suitable for analysing large numbers of DNA samples or for forensic work. Polymorphisms that can be typed by PCR would be ideal but only a few are known: a GATA repeat (Roewer et al., 1992), three CA repeats (Mathias et al., 1994), a point mutation (Seielstad et al., 1994), the presence or absence of *YRRM2* sequences (Nakahori et al., 1994) and a minisatellite which is being studied by

Correspondence to: Dr. F.R. Santos, Universidade Federal de Minas Gerais, Departamento de Bioquimica, Av. Antonio Carlos 6627, Caixa Postal 486, 31.270-910 Belo Horizonte MG, Brazil. Tel. (55-31) 441-5611; Fax (55-31) 441-5963; e-mail: fabricio@oraculo.lcc.ufmg.br or Dr. C. Tyler-Smith, CRC Chromosome Molecular Biology Group, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. Tel. (44-1865) 275-222; Fax (44-1865) 275-283; e-mail: chris@bioch.ox.ac.uk

Abbreviations: alphoid, human tandemly repeated DNA sequence with approx. 170-bp periodicity; αh, alphoid locus; Boleth, a human 46,XY cell line; bp, base pair(s); BSA, bovine serum albumin; CEPH, Centre d'Etude du Polymorphisme Humain; dNTP, deoxyribonucleoside triphosphate; DTT, dithiothreitol; h, pair of heteroduplex bands; kb, kilobase(s) or 1000 bp; MDE, mutation detection enhancement; MVR-PCR, minisatellite variant repeat-PCR; nt, nucleotide(s); OXEN, a human 49, XYYYY cell line; PCR, polymerase chain reaction; RFLP restriction-fragment length polymorphism; TAE, 40 mM Tris·acetate/1 mM EDTA pH 7.2; TBE, 45 mM Tris borate/1 mM EDTA pH 7.4; u, unit(s); YAC, yeast artificial chromosome; *YRRM2*, Y-located RNA recognition motif gene 2.

SSDI 0378-1119(95)00501-3

MVR-PCR (Jobling et al., 1994). More PCR polymorphisms are needed.

Alphoid satellite DNA is a tandemly repeated approx. 170-bp sequence found at all human centromeres (Choo et al., 1991). On the Y chromosome it shows both long-range variation due to different numbers of approx. 5.7-kb higher-order units per array, and short-range variation due to different numbers of approx. 170-bp subunits per unit (Tyler-Smith and Brown, 1987). All known human Y chromosomes contain approx. 5.7-kb units while some, in addition, contain a small number of approx. 6.0-kb units which have two additional approx. 170-bp subunits (Tyler-Smith and Brown, 1987). Since short-range polymorphisms of alphoid DNA have been successfully analysed by PCR on several autosomes (Warburton et al., 1991), we have investigated Y alphoid DNA polymorphisms using PCR. We find that the polymorphisms are more complex than those reported on the autosomes, and that they provide a rich source of new Y variants.

RESULTS AND DISCUSSION

(a) PCR amplification of variant Y alphoid units

We set out to establish a PCR-based assay which would distinguish between the 5.7-kb alphoid units present on all Y chromosomes and the 6.0-kb units detected only on some Y chromosomes. PCR primers were therefore chosen from within the 342-bp region that is present in 6.0-kb Y alphoid units but absent from the 5.7-kb units (Tyler-Smith and Brown, 1987). As expected, they produced a single approx. 285-bp product from a cloned 6.0-kb unit (Fig. 1, lane 2) and no product from a cloned 5.7-kb unit (lane 3). Similarly, no amplification was seen from female human DNA (lane 6), but the approx. 285-bp and additional bands were seen when male DNA was used (lanes 4 and 5). Work described in section **e** shows that the approx. 285-bp band contains a mixture of fragments with sizes between 281 bp and 285 bp; for simplicity we refer to the band as 'approx. 285-bp'. Surprisingly, the approx. 285-bp band was amplified both from males who have 6.0-kb units (lane 5) and males thought to lack 6.0-kb units (lane 4). The faint large (>662 bp) and small (<90 bp) bands vary in intensity in different experiments. The large band may be single-stranded DNA and the smaller band may represent primer dimers, while we show in section **d** that the prominent intermediate band migrating just behind the approx. 285-bp fragment represents heteroduplex molecules.

(b) Genomic localisation of the PCR products

In order to understand the generation of the approx. 285-bp band from males lacking 6.0-kb units, we investigated the genomic origins of the approx. 285-bp products. PCR amplification was carried out on two sets of YAC clones from known positions within Y alphoid arrays which either lack (Boleth, the source of the CEPH library; Albertsen et al., 1990) or contain (OXEN; Neil et al., 1990; Cooper et al., 1993; Foote et al., 1992) 6.0-kb units. All four YACs from the CEPH library produced the approx. 285-bp fragment (Fig. 2b). Since these YACs fall into two separate contigs (Fig. 2a), there must be at least two copies of the target locus. In contrast, only four of the nine OXEN YACs produced the approx. 285-bp fragment (Fig. 2b). The YACs which give the product again fall into two separate contigs (Fig. 2a). Thus the approx. 285-bp products arise from at least two target loci, one towards each edge of the alphoid array (Fig. 2a).

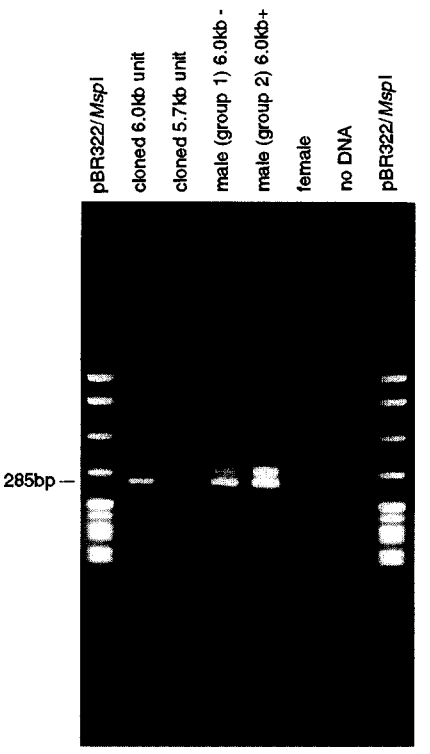


Fig. 1. Amplification of variant alphoid sequences from cloned alphoid units and human genomic DNA. Samples were run on a 4% (w/v) Nusieve/agarose (3:1) gel in 0.5 × TAE buffer containing ethidium bromide. Samples are identified at the top of each lane. **Methods:** PCR primers were U972 5′-TCTGAGACACTTCTTTGTGGTA and L1214 5′-CGCTCAAAATATCCACTTTCAC (Tyler-Smith and Brown, 1987). Reactions (Saiki et al., 1988) were carried out in a volume of 50 μl containing 40 ng DNA/1.25 u *Taq* polymerase (Promega, Madison, WI, USA)/1 × *Taq* buffer (Promega)/1.5 mM $MgCl_2$/250 μM dNTPs/1 μM of each primer using a Perkin Elmer Cetus or MJ Research thermal cycler. Conditions were 94°C for 3 min followed by 30 cycles of 94°C for 30 s, 65°C for 30 s and 72°C for 1 min.

(c) Detection of a *Hin*dIII site polymorphism

We wished to know how similar the four cloned copies of the target locus were and their relationship to the 6.0-kb units. Each of the PCR products was therefore

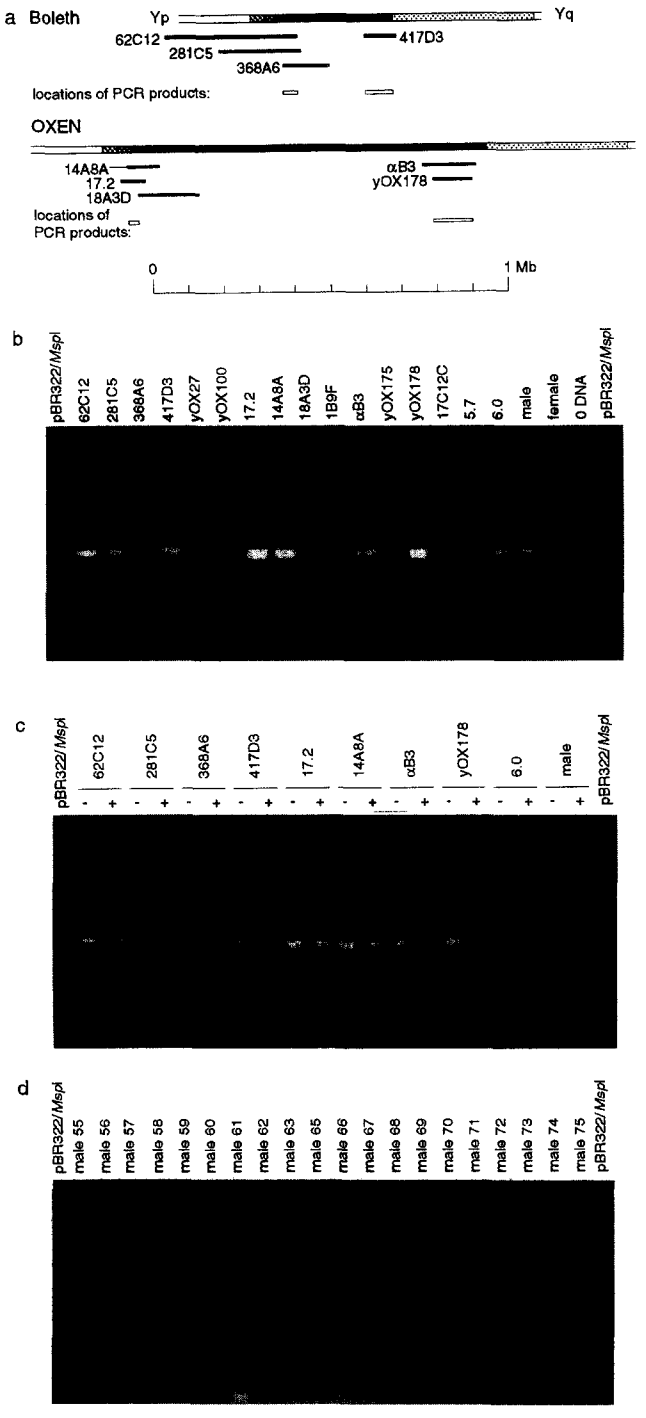


Fig. 2. Analysis of variant alphoid sequences. (**a**) Location of the alphoid YAC clones which contained the approx. 285-bp product on the Boleth (upper) or OXEN (lower) Y chromosome (Cooper et al., 1993). YAC clones containing Y alphoid DNA were derived from several libraries (Neil et al., 1990; Albertsen et al., 1990; Foote et al., 1992; Cooper et al., 1993) and most have been characterised in detail (Neil et al., 1990; Cooper et al., 1993). 368A6 (Albertsen et al., 1990) and yOX178 (Foote et al., 1992) were localised on their chromosomes of origin by restriction site mapping as previously described (Cooper et al., 1993). 17C12C is an OXEN-derived YAC lacking alphoid DNA and was used as a negative control. The additional OXEN YACs could not be placed precisely either because they were located in the large homogeneous central region (1B9F) or because they were extensively rearranged (yOX27, yOX100, yOX175). (**b**) PCR products obtained from the individual clones identified at the top of each lane. (**c**) Undigested (−) or *Hin*dIII-digested (+) samples from (**b**). 10 μl of the PCR reaction were digested with *Hin*dIII (Boehringer-Mannheim). (**d**) PCR products from males 55–75 after digestion with *Hin*dIII. (**b–d**) Ethidium bromide-stained Nusieve/agarose gels, but (**b**) and (**c**) were run for 6 h, while (**d**) was run for 3 h.

characterised by digestion with *Hph*I, *Sau*96I, *Hgi*AI, *Asp*700 and *Hin*dIII. All products showed a similar cleavage pattern with *Hph*I, *Sau*96I and *Hgi*AI. The products from the right-hand edge of both arrays contained an *Asp*700 site while the left-hand edge products did not (results not shown). The products from the OXEN right hand edge contained a *Hin*dIII site (Fig. 2c), but no *Hin*dIII cleavage was detected in the Boleth right hand edge product or in either left-hand edge product. This explains why 6.0-kb units are detected in *Hin*dIII digests only in a subset of males despite the presence of the PCR target region in all males. It is not known what size of units carry the target region at places where the *Hin*dIII site is absent.

Since this *Hin*dIII site polymorphism is expected to correspond to the previously described presence or absence of 6.0-kb units detected in *Hin*dIII digests, PCR products from a panel of 93 males were tested for the presence or absence of the *Hin*dIII site (e.g., Fig. 2d). Some individuals (e.g., male 55) showed no cleavage with *Hin*dIII. In others a small proportion (male 58) or about half (e.g., male 59) of the PCR product was cleaved. In 92 of the 93 males there was agreement between the PCR analysis and the previous hybridisation result (Table I): 53 males were +, +; 39 males were −, −. In one individual (male 60) the 4.1-kb and 1.9-kb alphoid *Hin*dIII fragments had been detected by hybridisation but no cleavage of the PCR product was seen (Fig. 2d). Sample mix-up and failure of the restriction enzyme to digest were excluded as explanations for this, but a more thorough analysis was not possible because of the limited amount of DNA available from this individual. Despite this anomaly, the *Hin*dIII site variants can conveniently be detected in genomic DNA using PCR. Indeed, since genomic DNA hybridisation results are complicated by the presence of 4.1-kb units, the PCR assay is the method of choice for detecting 6.0-kb units.

(d) Detection of polymorphic heteroduplex patterns

PCR products derived from genomic DNA, but not from individual clones, generally showed an intermediate band migrating slightly more slowly than the approx. 285-bp fragments on agarose gels (Figs. 1 and 2). On standard polyacrylamide gels this band could be resolved into more complex patterns (not shown) but in most cases the best resolution was obtained on MDE-polyacryl-

TABLE I

PCR haplotypes of 93 Y chromosomes

Ind. no.[a]	Bg.[b]	Gp.[c]	*Hin*dIII site[d]		Heteroduplex bands[e]												Heteroduplex haplotype[f]
			hyb	PCR	h1	h2	h3	h4	h5	h6	h7	h8	h9	h10	h11	h12	
1	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
2	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
3	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
4	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
5	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
6	C	4	+	+	-	-	-	-	+	+	-	-	-	-	+	+	XIII
7	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
8	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
9	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
10	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
11	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
12	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
13	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
14	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
15	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
16	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
17	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
18	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
19	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
20	C	1	-	-	-	+	-	-	-	-	-	-	-	-	-	-	VI
21	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
22	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
23	C	2	+	+	-	-	-	+	+	-	-	-	-	-	-	-	VIII
24	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
25	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
26	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
27	M	2	+	+	-	-	-	-	-	-	-	-	-	-	-	-	I
28	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
29	C	3	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
30	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
31	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
32	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
33	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
34	C	3	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
35	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
36	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
37	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
38	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
39	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
40	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
41	U	4	+	+	+	-	-	-	+	+	-	-	-	-	-	-	V
42	U	3	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
43	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
44	O	4	+	+	-	-	-	-	+	+	-	-	-	-	+	-	XII
45	O	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
46	O	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
47	N	-	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
48	N	-	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
49	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
50	U	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
51	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
52	C	3	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
53	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
54	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
55	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
56	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
57	C	1	-	-	-	-	+	-	-	-	-	-	-	-	-	-	VII
58	C	2	+	+	+	-	-	-	+	+	-	-	-	-	-	-	V
59	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
60	C	2	+	-	+	-	-	-	-	-	-	-	-	-	-	-	II
61	C	4	+	+	-	-	-	-	-	-	-	-	-	-	-	-	I
62	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
63	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
65	M	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
66	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
67	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
68	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
69	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
70	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
71	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
72	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
73	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
74	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
75	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
76	N	4	+	+	+	-	-	-	+	+	+	-	+	-	-	-	X
77	N	4	+	+	+	-	-	-	+	+	-	-	-	-	-	-	V
78	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
79	N	4	+	+	+	-	-	-	+	+	+	+	-	-	-	-	IX
80	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
81	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
82	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I
83	C	2	+	+	-	-	-	-	+	-	-	-	-	-	-	-	IV
84	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
85	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
86	N	4	+	+	+	-	-	-	+	+	-	-	-	-	-	-	V
87	O	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
88	O	2	+	+	-	-	-	-	-	-	-	-	-	-	-	-	I
89	M	2	+	+	+	-	-	-	+	-	-	-	-	+	-	-	XI
90	A	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
93	A	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
94	C	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III
95	C	1	-	-	+	-	-	-	-	-	-	-	-	-	-	-	II
96	U	2	+	+	+	-	-	-	+	-	-	-	-	-	-	-	III

amide gels (Fig. 3). The origin of the intermediate bands was investigated by mixing the products from different clones (Fig. 3a). The PCR product from the left (lane 2) or right (lane 3) edge of the Boleth array did not show such a band (see Fig. 2a for the locations of the PCR products). Simply mixing the two products did not lead to the formation of additional bands (lane 4), but mixing followed by denaturation and annealing produced two new bands (lane 5) which correspond to bands in the Boleth genomic DNA sample (lane 6). Similar experiments with PCR products from the OXEN array also led to the formation of two bands (lane 10) which are different from the Boleth bands but correspond to two of the four bands in the OXEN genomic sample (lane 11). The other pair of OXEN bands is likely to be derived from a more distal right edge locus which is absent from the YAC clones available. This experiment shows that the intermediate bands are composed of heteroduplex pairs formed from left and right edge products and that they are polymorphic on different Y chromosomes.

We therefore examined the heteroduplex patterns in a panel of 93 Y chromosomes derived from a wide range of human backgrounds (Table I). 89 out of the 93 showed a heteroduplex pattern; individual patterns contained between one and five pairs of bands (Fig. 3b). There were 12 different pairs of heteroduplex bands (designated h1–h12: Fig. 3c). Each pattern provides a haplotype for the Y chromosome. In all, 13 haplotypes (designated I–XIII) could be distinguished (Fig. 3b, Table I). Haplotypes III and VIII were more clearly distinguished on a standard polyacrylamide gel (results not shown). The four individuals who did not show a heteroduplex pattern produced one when annealed with the PCR product from the left edge YAC 62C12 (results not shown), suggesting that they carry a deletion of this locus or a mutation which prevents priming.

[a] Ind no., individual number.

[b] Bg, background (A, Amerindian; C, Caucasian; M, Mongoloid; N, Negroid; O, Oceanic; U, unknown).

[c] Gp, previously determined Y chromosome group. Origin of samples: Human DNA samples included 90 of the 91 samples described previously (Mathias et al., 1994), derived from a wide range of human populations. Additional samples were males Ind. no. 94, Caucasian; Ind. no. 95, Caucasian (Boleth), the cell line used as the source of DNA for the CEPH YAC library (Albertsen et al., 1990); Ind. no. 96, origin unknown.

[d] The presence (+) or absence (−) of the *Hin*dIII site was determined either by hybridisation (hyb) or PCR (PCR).

[e] The heteroduplex bands revealed in each DNA sample were determined as shown in Fig. 3b and c.

[f] The heteroduplex haplotype I–XIII is the combination of heteroduplex bands in each sample. Examples of each haplotype are shown in Fig. 3b and c.

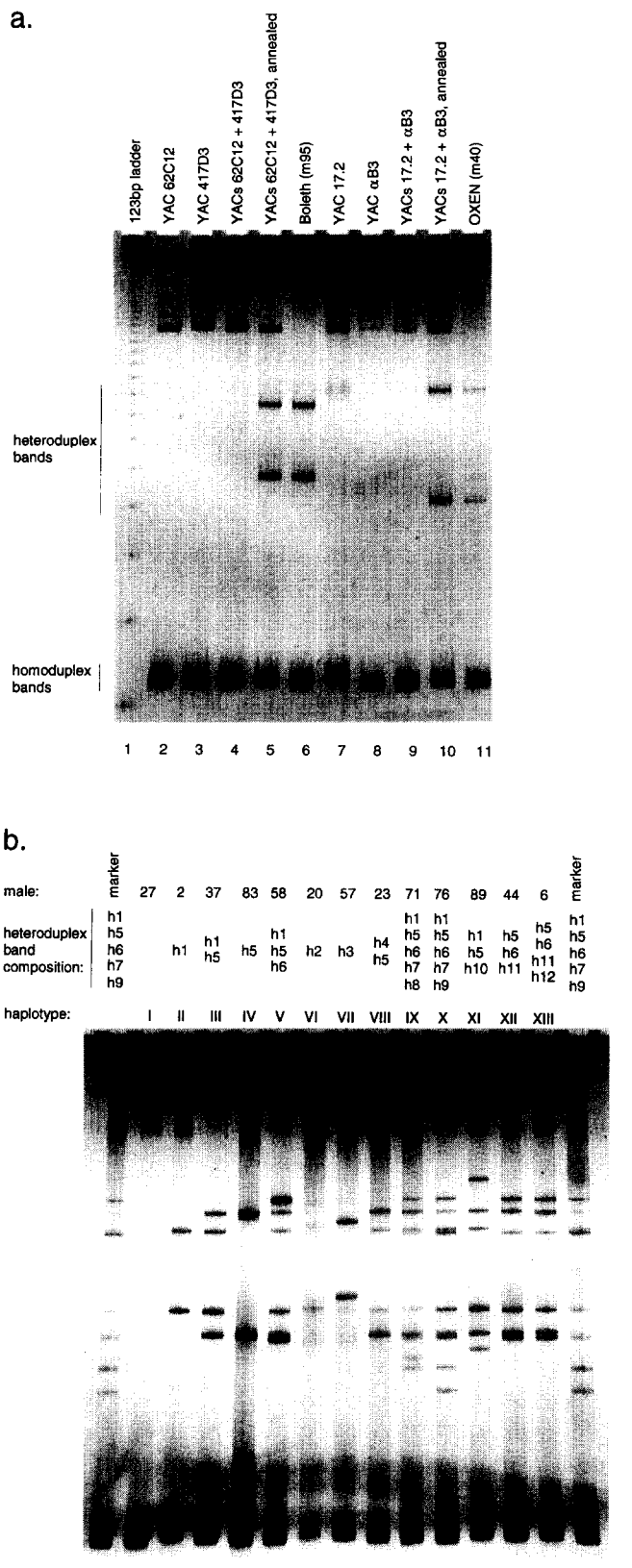
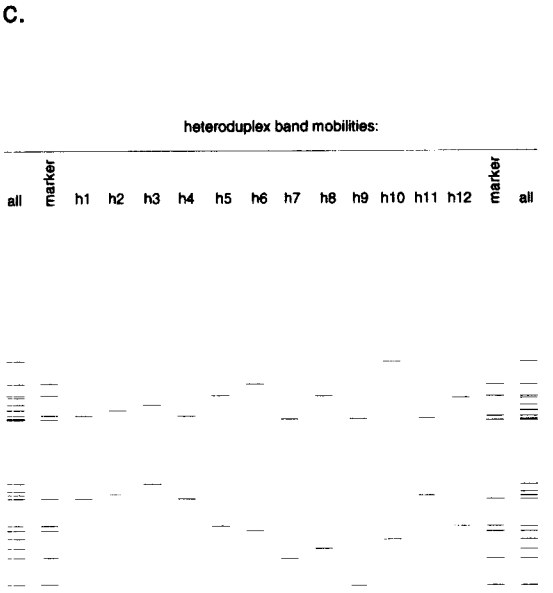


Fig. 3. Detection of heteroduplex bands. (**a**) Formation of heteroduplex bands from PCR products. Samples are identified at the top of each lane. **Methods:** 5 μl of the amplified product were loaded on a 1 × MDE-polyacrylamide gel (AT Biochem, Malvern, PA, USA) in 1 × TBE buffer and run at 100 V for 15 h. In mixing experiments, 5 μl samples of the amplified products were mixed, denatured by boiling for 30 s and allowed to cool to room temperature to form heteroduplexes. Bands were visualised by silver staining as described previously (Santos et al., 1993b). (**b**) Heteroduplex band patterns from selected males on silver-stained 1 × MDE-polyacrylamide gel. The heteroduplex markers are an annealed mixture of PCR products from clones containing αhL, αh1, αh5, αh6, αh7 and αh9. (**c**) Schematic representation of individual heteroduplex band pairs making up the patterns in (**b**). **Nomenclature:** the nomenclature used is as follows: Loci and alleles are designated αhL for the left-hand locus and αh1 to αh12 for the right-hand loci. The combination of αhL with each right-hand locus produces the corresponding pair of heteroduplex bands h1 to h12: for example, αhL + αh1 gives h1. The set of 0, 1 or more heteroduplex pairs found in an individual defines the heteroduplex haplotype I to XIII.



(e) Sequence basis of the heteroduplex patterns

We wished to address two questions raised by the heteroduplex patterns. Firstly, why do the heteroduplexes show such large mobility shifts compared with the homoduplexes? Secondly, what is the structural basis for the polymorphisms seen on different Y chromosomes? PCR products from YAC clones and selected individuals were therefore cloned and sequenced (Fig. 4). This procedure can produce artifacts due to misincorporation by the *Taq* polymerase. Two checks were used to exclude clones containing errors: the PCR product derived from the clone was tested in mixing experiments of the kind used in Fig. 3a and shown to produce heteroduplex bands corresponding to the original source, and sequence information was only used if two or more independent clones showed the same sequence.

The results of the sequencing are summarised in Fig. 4. Single bp variants were detected at nine positions and one 4-bp insertion/deletion was seen. We will refer to the 4-bp difference as a deletion because it lies in an alphoid subunit of 167-bp, while standard alphoid subunits are 170–171-bp long. The left edge product differed from all others examined by the 4-bp deletion at nt 169–172 and

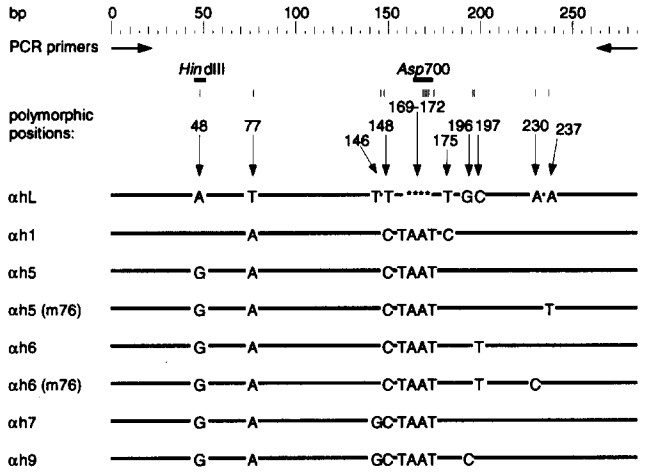


Fig. 4. Sequence variants detected in cloned PCR products. Variants were detected at ten positions (nine point mutations, one 4-nt deletion) as indicated. For αhL, the base at each of these positions is shown. For αh1 to αh9, the sequence is shown only when it differs from αhL. Methods are available from C.T.-S. upon request. The number of sequence determinations from independent clones used for each heteroduplex locus was: αhL: YAC 368A6 (4), YAC 17.2 (1), male 1 (1), male 6 (3), male 76 (1). αh1: YAC 417D3 (3), male 1 (1), male 76 (1). αh5: YAC αB3 (3), male 6 (3). αh5(m76): male 76 (2). αh6: plasmid pYαI (3). αh6(m76): male 76 (3). αh7: male 76 (5). αh9: male 76 (3). These sequences have been submitted to GenBank and have been assigned the accession Nos. U16679 (αhL), U16675 (αh1), U16676 (αh5), U16677 (αh5(m76)), U16682 (αh6), U16678 (αh6(m76)), U16680 (αh7) and U16681 (αh9).

two point mutations at nt 77 and 148. The deletion is likely to account for the large mobility shift of the heteroduplexes, as observed in other systems (Van den Akker et al., 1992). Additional point mutations at nt 48, 146, 175, 196, 197, 230 and 237 distinguished clones able to form different heteroduplex bands in reconstruction experiments, although in two cases (h5, h5(m76); h6, h6(m76)) sequences differing at a single position produced heteroduplexes with indistinguishable mobility. This is probably because a mismatch near the end of a heteroduplex has little effect on its mobility. We therefore conclude that polymorphic heteroduplex patterns are seen because individual Y chromosomes carry a left edge copy of the target locus with a deletion and a variable number of loci which lack the deletion and differ by several point mutations.

The sequencing results also account for the differences in restriction pattern described above. The 4-bp deletion destroys an *Asp*700 site which is therefore absent in the left edge products and present in the right edge products. The A/G point mutation at position 48 changes AAACTT present in the left edge product and the Boleth right edge product (generating h1) to AAGCTT, a *Hin*dIII site present in all of the other right edge products sequenced. These correspondences confirm the reliability of the sequences determined from cloned PCR products. *Hin*dIII sites are absent from the PCR product when only heteroduplex bands h1, h2 or h3 are present (haplotypes II, VI, VII) but are present when additional heteroduplex bands are present (haplotypes III–V, VII–XIII).

(f) Evolutionary origin of the polymorphisms

The target loci are present in a small proportion of Y alphoid units situated near both ends of the arrays (Fig. 2). One explanation for this distribution is as follows: The array would initially have contained the target region in all alphoid units. 5.7-kb units lacking the region arose by a deletion event in the middle of the array and spread through most of it. Only the edges of the array, which are expected to be homogenised least extensively by expansion and contraction (Smith, 1976), and a few islands within the array now retain the target region. In 4/93 chromosomes the left edge target region has also been lost; Y chromosomes which have lost all copies of the region may exist, although they have not yet been detected. Target loci from the left and right edges no longer exchange sequence information with one another and so have accumulated differences giving rise to the heteroduplex patterns. Alternative evolutionary scenarios are also possible.

(g) Use of the alphoid PCR haplotypes

Informative PCR haplotypes for the Y chromosome can in principle be constructed using the existing microsatellite markers (Roewer et al., 1992; Mathias et al., 1994). Although such haplotypes could distinguish different Y chromosomes, chromosomes with the same haplotype will not necessarily be closely related to one another because a particular allele may arise independently. Thus microsatellite haplotypes are of limited use for evolutionary purposes (Mathias et al., 1994). In contrast, differences between the alphoid PCR haplotypes described here are largely due to point mutations and deletion events and consequently show a lower level of homoplasy. We have therefore compared the alphoid PCR haplotypes with previous compound haplotypes of the same chromosomes which resulted in the identification of seven groups: 1, 2, 3, 4, Biaka Pygmies, !Kung and Amerindians (Mathias et al., 1994) in order to evaluate their likely usefulness in evolutionary studies (Table II).

Haplotype I, the absence of heteroduplex bands, is seen in three different groups: 2, 4 and the !Kung. It does not imply a similarity between these Y chromosomes; indeed, addition of the left-hand edge product to the !Kung individual generated a novel pattern which emphasises the distinctness of this chromosome (results not shown). Haplotype II was both the most common in the sample (37/93, 40%) and the most widespread, occurring in four

TABLE II

Frequency of heteroduplex haplotypes I–XIII in Y chromosome groups

Group[a]	Heteroduplex haplotype[b]													Total
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	
1		30[c]				1	1							32
2	2	1	25	1	1			1			1			32
3		4												4
4	1				3				13	1		1	1	20
Pygmies			2											2
!Kung	1													1
Amerindians		2												2
Total	4	37	27	1	4	1	1	1	13	1	1	1	1	93

[a] Previously defined group to which Y chromosome belongs (Mathias et al., 1994). The most frequent groups were numbered 1 to 4; less frequent groups were named after the population in which they were found (Pygmies, !Kung, Amerindians).

[b] Selected heteroduplex haplotypes are illustrated in Fig. 3 and the haplotype of each individual is shown in Table I.

[c] The number of chromosomes from each group carrying each heteroduplex haplotype is calculated from the data in Table I. The total number of chromosomes in each group and the total with each haplotype are also shown.

groups. It accounted for 30/32 (94%) of group-1 members; the other two group-1 individuals (haplotypes VI and VII) had simple variants of haplotype II. It was also found in all group-3 chromosomes examined (4/4) and both Amerindian chromosomes. The occurrence of haplotype II in one member of group 2 (m60, the individual who gave different results for the *Hin*dIII site in the hybridisation and PCR assays) is not understood at present. Haplotype III is found commonly in group-2 chromosomes (25/32, 78%) and also in the two Biaka Pygmies. Some subtle differences were noted in the haplotype-III patterns. In some individuals (e.g., m40 in Fig. 3a) the h1 bands were fainter than the h5 bands, while in others (e.g., m37 in Fig. 4) the two pairs of bands were of similar intensity. All of the Caucasian carriers of haplotype III and two Mongoloids (m36 and m63) were of the former type, while the remaining Mongoloid and Biaka Pygmy carriers were of the latter type. Haplotypes IV and VIII can be derived from III by single changes (loss of heteroduplex h1; replacement of h1 by h4, respectively) and are also found in group 2. Haplotype V is found in both group-2 and group-4 chromosomes. This could be because it was the haplotype of the common ancestor of the two groups; alternatively, it could have arisen twice, or different sequences might give rise to heteroduplexes with the same mobility. Haplotype IX is common in group-4 chromosomes (13/20, 65%) and haplotype X in group 4 is a variant of it. Haplotype XI is found in a single group-2 chromosome and the final two haplotypes, XII and XIII are found in single members of group 4.

Thus the alphoid PCR haplotypes show striking correspondences to the previous groups but also reveal heterogeneity within groups (especially groups 2 and 4) and help us to understand the relationships between different groups. In particular, a close relationship between group-1 Y chromosomes (common in Europeans), group-3 Y chromosomes (common in Indians; Jobling, 1994) and Amerindian Y chromosomes is seen, in agreement with previous observations (Mathias et al., 1994). It seems likely that haplotype II was present in the common ancestor of the three groups. Group-2 Y chromosomes are widespread in many non-African populations (Jobling, 1994; Mathias et al., 1994). The alphoid heteroduplex haplotypes suggest that the Biaka Pygmy Y chromosomes are the most closely related African Y chromosomes so far characterised.

This haplotyping system is suitable for the analysis of large numbers of samples and may be useful for population surveys. Results from different laboratories can be compared using markers derived from clones (Fig. 4) or a set of standard males. The system can also be used when limited amounts of low-molecular-weight DNA are available. In forensic cases information about the possible ethnic origin of a sample may be obtained, but such results must be interpreted with extreme caution until more population data are available and will always be of limited use because of the occurrence of admixture. It should be possible to develop similar haplotyping systems for other Y chromosome repeats, if necessary using artificially constructed 'heteroduplex generator' molecules containing an insertion or deletion. With a series of such assays, the polymorphism of the Y chromosome can be analysed in detail using PCR.

(h) Conclusions

(*1*) We have developed a PCR-based system for haplotyping the human Y chromosome.

(2) The system is unusual in that it is both highly polymorphic and useful for evolutionary studies. This is possible because it detects a small number of target loci which differ by point mutations and thus individually evolve slowly, but are present or absent in varied combinations on different Y chromosomes so that the pattern as a whole evolves more rapidly.

ACKNOWLEDGEMENTS

We thank the original donors and Neal Mathias for providing DNA samples and Ed Southern, William Brown, Neal Mathias and Mark Jobling for encouragement and comments on the manuscript. This work was initiated during a UNESCO-funded visit of F.R.S. to Oxford. F.R.S. and S.D.J.P. are supported by the Conselho Nacional de Pesquisa; C.T.-S. is supported by the Cancer Research Campaign.

REFERENCES

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., Le Paslier, D. and Cohen, D.: Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. Proc. Natl. Acad. Sci. USA 87 (1990) 4256–4260.

Casanova, M., Leroy, P., Boucekkine, C., Weissenbach, J., Bishop, C., Fellous, M., Purrello, M., Fiori, G. and Siniscalco, M.: A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230 (1985) 1403–1406.

Choo, K.H., Vissel, B., Nagy, A., Earle, E. and Kalitsis, P.: A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. 19 (1991) 1179–1182.

Cooper, K.F., Fisher, R.B. and Tyler-Smith, C.: Structure of the sequences adjacent to the centromeric alphoid satellite DNA array on the human Y chromosome. J. Mol. Biol. 230 (1993) 787–799.

Foote, S. Vollrath, D., Hilton, A. and Page, D.: The human Y chromosome: overlapping DNA clones spanning the euchromatic region. Science 258 (1992) 60–66.

Jakubiczka, S., Arnemann, J., Cooke, H.J., Krawczak, M. and Schmidtke, J.: A search for restriction fragment length polymorphism on the human Y chromosome. Hum. Genet. 84 (1989) 86–88.

Jobling, M.A.: A survey of long-range DNA polymorphisms on the human Y chromosome. Hum. Mol. Genet. 3 (1994) 107–114.

Jobling, M.A., Fretwell, N., Dover, G.A. and Jeffreys, A.J.: Digital coding of human Y chromosomes: MVR-PCR at Y-specific minisatellites. Cytogenet. Cell Genet. 67 (1994) 390.

Lucotte, G. and Ngo, N.Y.: p49f, a highly polymorphic probe, that detects TaqI RFLPs on the human Y chromosome. Nucleic Acids Res. 13 (1985) 8285.

Malaspina, P., Persichetti, F., Novelletto, A., Iodice, C., Terrenato, L., Wolfe, J., Ferraro, M. and Prantera, G.: The human Y chromosome shows a low level of DNA polymorphism. Ann. Hum. Genet. 54 (1990) 297–305.

Mathias, N., Bayés, M. and Tyler-Smith, C.: Highly informative compound haplotypes for the human Y chromosome. Hum. Mol. Genet. 3 (1994) 115–123.

Nakahori, Y., Kobayashi, K., Komaki, R., Matsushita, I. and Nakagome, Y.: A locus of the candidate gene family for azoospermia factor (YRRM2) is polymorphic with a null allele in Japanese males. Hum. Mol. Genet. 3 (1994) 1709.

Neil, D.L., Villasante, A., Fisher, R.B., Vetrie, D., Cox, B. and Tyler-Smith, C.: Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nucleic Acids Res. 18 (1990) 1421–1428.

Oakey, R. and Tyler-Smith, C.: Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. Genomics 7 (1990) 325–330.

Roewer, L., Arnemann, J., Spurr, N.K., Grzeschik, K.-H. and Epplen, J.T.: Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. 89 (1992) 389–394.

Saiki, R., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A.: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 (1988) 487–491.

Santos, F.R., Epplen, J.T. and Pena, S.D.J.: Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism. In: Pena, S.D.J., Chakraborty, R., Epplen, J.T. and Jeffreys, A.J. (Eds.), DNA Fingerprinting: State of the Science. Birkhäuser, Basel, 1993a, pp. 261–265.

Santos, F.R., Pena, S.D.J. and Epplen, J.T.: Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum. Genet. 90 (1993b) 655–656.

Seielstad, M.T., Hebert, J.M., Lin, A.A., Underhill, P.A., Ibrahim, M., Vollrath, D. and Cavalli-Sforza, L.L.: Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum. Mol. Genet. 3 (1994) 2159–2161.

Smith, G.P.: Evolution of repeated DNA sequences by unequal crossover. Science 191 (1976) 528–535.

Spurdle, A.B. and Jenkins, T.: The Y chromosome as a tool for studying human evolution. Curr. Opin. Genet. Dev. 2 (1992a) 487–491.

Spurdle, A. and Jenkins, T.: The search for Y chromosome polymorphism is extended to negroids. Hum. Mol. Genet. 1 (1992b) 169–170.

Tyler-Smith, C. and Brown, W.R.A.: Structure of the major block of alphoid satellite DNA on the human Y chromosome. J. Mol. Biol. 195 (1987) 457–470.

Van den Akker, K., Braun, J.E.F., Pals, G., Lafleur, M.V.M. and Retèl, J.: Single base mutations can be unequivocally and rapidly detected by analysis of DNA heteroduplexes, obtained with deletion-mutant instead of wild-type DNA. Nucleic Acids Res. 20 (1992) 6745–6746.

Warburton, P.E., Grieg, G.M., Haaf, T. and Willard, H.F.: PCR amplification of chromosome-specific alpha satellite DNA: definition of centromeric STS markers and polymorphic analysis. Genomics 11 (1991) 324–333.