

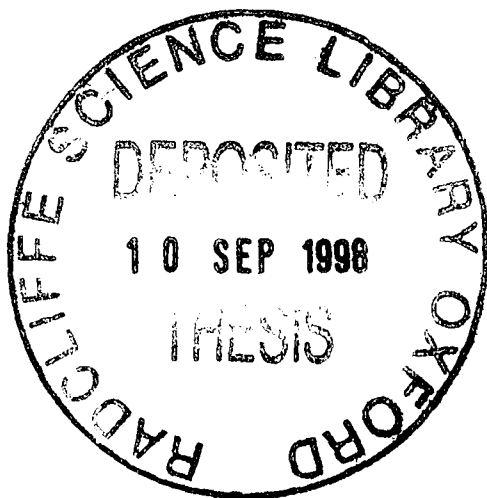
**Mapping studies of the centromeric region of the
human Y chromosome**

By

Gareth Williams

**Thesis submitted for the degree of Doctor of
Philosophy at the University of Oxford**

**St Catherine's College
Hilary term 1998**



Abstract

Gareth Williams
St Catherine's College

Submitted for the degree
of Doctor of Philosophy
Hilary term 1998

Mapping studies of the centromeric region of the human Y chromosome

Construction of a map of a human centromeric region is very important in order to understand the organisation of this essential part of the chromosome. A YAC contig map has been assembled of the pericentric 10 Mb of the human Y chromosome, giving coverage of Yp from the large X-Y homologous region through to the alphoid satellite of the centromere, and from the alphoid DNA to the proximal unique sequences on Yq. The Yp map has one remaining gap between TSPY1 and the *AMELY* region, while two gaps separate the satellite region on Yq from the other two contigs.

After constructing the map, the known genes were localised to the region. One Yq gene, *DFFRY*, was discounted as a potential anti-Turner syndrome gene by analysis of rearranged Y chromosomes.

Detection of a block of duplicated sequence on Yp led to the confirmation of the existence of an inversion polymorphism, which was then found to be correlated with a major subclass of sex-reversed individuals, who have X-Y chromosomal breakpoints within the inverted region.

These results not only give a far more extensive and detailed map of this region than before, but also show that understanding the organisation of the region has important consequences for a number of genetic disorders.

Acknowledgements

I would like to thank Chris Tyler-Smith for all his assistance and support over the lifetime of this thesis. I am also grateful to Professor Southern for his help and encouragement. Thanks are also due to the Medical Research Council, for providing me with funding; and to the MRC Human Genome Resource Centre in Hinxton for their many provisions toward human genome research, not the least of which was supplying me with yeast clones and YAC libraries. Many thanks also to Tasnim Alam, for assistance and chats over the last three years; as well as to William Brown and Martin Johnson for their help in the laboratory.

On a personal level, unlimited thanks must be extended to my family for all their support; and finally, I am unable to thank Catherine enough, for everything.

Contents

ABSTRACT	2
CHAPTER 1: INTRODUCTION.....	7
1.1: THE HUMAN GENOME PROJECT	7
1.2 THE HUMAN Y CHROMOSOME	11
1.2.1 <i>Y Chromosome evolution and function</i>	11
1.2.2 <i>Structural overview of the human Y chromosome</i>	14
1.2.3 <i>Mapping the Y chromosome</i>	18
1.2.4 <i>The Foote et al. map</i>	18
1.2.5 <i>Problems with this map</i>	19
1.3 THE CENTROMERE / KINETOCHORE COMPLEX	21
1.4 CENTROMERES OF <i>SACCHAROMYCES CEREVISIAE</i>	22
1.4.1 <i>Structural analysis</i>	22
1.4.2 <i>Functional analysis</i>	23
1.4.3 <i>Proteins</i>	23
1.5 CENTROMERES OF <i>SCHIZOSACCHAROMYCES POMBE</i>	25
1.6 CENTROMERES OF <i>DROSOPHILA MELANOGASTER</i>	27
1.7 CENTROMERES OF <i>HOMO SAPIENS</i>	29
1.7.1 <i>Structural analysis</i>	29
1.7.2 <i>Functional analysis</i>	30
1.7.3 <i>Centromeric proteins</i>	32
1.8 ANALYSIS OF CENTROMERES	34
1.8.1 <i>Why study centromeres?</i>	34
1.8.2 <i>Problems with centromeric analysis</i>	36
1.9 PURPOSE OF THIS PROJECT	37
CHAPTER 2: MATERIALS AND METHODS	41
2.1 GROWTH AND STORAGE OF BACTERIAL, YEAST, AND MAMMALIAN CELLS	41
2.1.1 <i>Bacterial Cells</i>	41
2.1.2 <i>Yeast Cells</i>	41
2.1.3 <i>Mammalian Cells</i>	42
2.2 PROBES USED IN THIS WORK	42
2.3 PREPARATION OF DNA	43
2.3.1 <i>Preparation of plasmid DNA</i>	43
2.3.2 <i>Preparation of PAC DNA</i>	44
2.3.3 <i>Preparation of YAC DNA</i>	44
2.4 MANIPULATION OF DNA	45
2.4.1 <i>Restriction endonuclease digestion</i>	45
2.5 GEL ELECTROPHORESIS	46
2.5.1 <i>Solutions and reagents</i>	46
2.5.2 <i>Conventional gels</i>	46
2.5.3 <i>Preparative gels</i>	46
2.5.4 <i>Purification of DNA from agarose gels</i>	47
2.5.5 <i>Pulsed-field gel electrophoresis</i>	47
2.5.6 <i>Gel viewing and photography</i>	47

2.6 SOUTHERN ANALYSIS.....	47
2.6.1 Blotting of gels.....	47
2.6.2 Hybridisation of filters.....	48
2.7 PLASMID END-RESCUE OF PROBES FROM YACs.....	50
2.7.1 Preparation of electrocompetent cells.....	50
2.7.2 Preparation of DNA.....	50
2.7.3 Electroporation.....	50
2.8 TRANSFORMATION OF YEAST CELLS	51
2.8.1 Reagents and solutions	51
2.8.2 Transformation of yeast	51
2.9 POLYMERASE CHAIN REACTION (PCR)	51
2.9.1 Reagents and solutions	51
2.9.2 Primers.....	52
2.9.3 PCR.....	53
2.10 FLUORESCENT IN-SITU HYBRIDISATION (FISH)	54
2.10.1 Preparation of metaphase spreads	54
2.10.2 Preparation of extended chromatin slides	54
2.10.3 Probe labelling	55
2.10.4 FISH.....	56
CHAPTER 3: MAPPING OF THE Y CHROMOSOME LONG ARM.....	58
3.1 CONTIGS 14A3C AND GMGY6.....	58
3.2 THE 68 BP SATELLITE CONTIG	70
CHAPTER 4: MAPPING OF THE Y CHROMOSOME SHORT ARM.....	75
4.1: THE ALPHOID CONTIG	75
4.2 THE TSPY CONTIG	83
4.3 MAPPING OF THE <i>AMELY</i> CONTIG	86
4.4: THE YAC 887G12	103
4.5: CLOSING THE REMAINING GAPS	112
CHAPTER 5: INVERSIONS AND DUPLICATIONS.....	114
5.1 A DUPLICATION ON THE SHORT ARM OF THE Y CHROMOSOME.....	114
5.2: AN X-Y HOMOLOGOUS REGION ON THE <i>AMELY</i> CONTIG.....	121
5.3: AN INVERSION POLYMORPHISM ON THE Y CHROMOSOME SHORT ARM.	124
CHAPTER 6: GENES AND MUTANTS.....	131
6.1 LOCALISING GENES ON THE CHROMOSOMAL MAP.....	131
6.2: DELETED CHROMOSOMES	141
6.3: A STUDY OF SEX-REVERSED INDIVIDUALS.	144
CHAPTER 7: DISCUSSION.....	155
7.1: PHYSICAL YAC MAPPING	155
7.1.1: 14A3C and GMGY6.....	155
7.1.2: The 68 bp satellite contig.....	156
7.1.3: The alphoid – TSPY contig	157
7.1.4: The <i>AMELY</i> contig.....	159
7.2: SEQUENCE FEATURES OF THE Y CHROMOSOME.....	161
7.3: THE EFFECT OF SEQUENCE ORGANISATION ON THE OCCURRENCE OF SEX REVERSAL.	166

7.4 IMPLICATIONS OF THESE RESULTS..... 168

7.4.1 *The Human Genome Project* 168

7.4.2 *Centromeres and MACs* 170

7.5 FURTHER WORK TO BE CARRIED OUT 172

REFERENCES 175

Chapter 1: Introduction

"Journey all over the universe in a map,
without the expense and fatigue of travelling,
without suffering the inconveniences
of heat, cold, hunger and thirst"

Cervantes, *Don Quixote*

1.1: The Human Genome Project

The Human Genome Project is the world-wide effort to map the entire human genome, eventually down to the sequence level. The human genome consists of some 3×10^9 base pairs, arranged in 23 chromosome pairs. An important part of the mapping project is to provide maps of each chromosome at varying resolutions. These maps are of a number of types – cytogenetic, genetic, or physical. All maps have the same basic form – a sequence of landmarks (physical or genetic), ordered with respect to one another along the chromosome; many (although not all) maps also show the distance between markers, either in physical terms (such as base pairs), or genetic distance, measured usually in centimorgans, cM. Typical landmarks are genes, restriction enzyme sites, repeated sequences such as Alu repeats, cytogenetic chromosome bands, chromosome deletion intervals (e.g., Vergnaud et al., 1986), or STSs – Sequence Tagged Sites, primer pairs which define a unique PCR product (Olson, Hood, Cantor & Botstein, 1989). This last type of landmark has been of great utility in rapidly assembling physical maps, as many STSs can be quickly and easily scored for presence or absence in many different samples.

Genetic maps show the relative locations along the chromosome of landmarks (not necessarily genes), and the distance between them in cM, a measure of the likelihood of recombination occurring between two markers when inherited. Such maps are very useful for a number of purposes; but have the major disadvantage that they can only be assembled using entire breeding organisms, as the markers must undergo recombination before the maps can be created. In addition, genetic linkage maps can only be assembled where polymorphisms exist amongst markers. The most common types of polymorphisms used are restriction fragment length polymorphisms (RFLPs – see Botstein et al., 1980), and variable number tandem repeats (such as microsatellites or minisatellites – see Jeffreys, Wilson & Thein, 1985; Weber 1990). One example of such a map is given in Donis-Keller et al. (1987). Polymorphisms are

used to trace inheritance patterns amongst family members, so as to obtain the most accurate data possible about the recombination distance between the markers under study; and to distinguish the two copies of each chromosome carried by an individual. For this reason, and for use in general population and genealogy studies, it is very useful to identify novel polymorphisms amongst populations. Genetic maps also cannot be assembled for certain regions of the genome – for many organisms, some part of the sex chromosomes does not undergo genetic recombination, and so cannot be mapped in this way. In mammals, the bulk of the Y chromosome does not recombine with the X chromosome, so genetic maps cannot be assembled for this region.

Physical maps, on the other hand, show the actual makeup of the genome, rather than the more abstract concept of ‘genetic distance’ as genetic maps do. The lowest resolution physical map is the cytogenetic map – certain staining techniques produce banding patterns unique to each chromosome, by means of which the chromosomes can be identified and distinguished from one another. Further landmarks may then be assigned to specific bands of the chromosome by means of *in situ* hybridisation. The resolution with which metaphase bands can be distinguished in this way is on the order of 10 Mb.

Another physical mapping method whereby markers can be assigned to specific intervals is chromosome deletion mapping. Panels of cell lines are assembled which have undergone deletions of the chromosome of interest to varying extents (these may be naturally occurring deletions, or artificially induced). The landmarks can then be located on the chromosome in deletion intervals determined from the panel of cell lines. Tyler-Smith et al. (1993) used this technique to localise the centromeric region of the Y chromosome from rearranged chromosomes.

A similar technique is radiation hybrid mapping, involving creating a panel of radiation hybrid cell lines. The genomic DNA to be studied is irradiated, thereby fragmenting it, and fused with a host cell line, such as hamster cells. The host cells integrate one or more of the genomic fragments into the genome, and can be stably propagated. Analysis of marker content can then be undertaken, the theory being that two closely-linked markers will co-localise to the same genomic fragment in several

clones. A number of radiation hybrid panels are widely available (see, for instance, Gyapay et al., 1996).

Higher resolution physical maps can be assembled over a small region of the genome by means of high-resolution fluorescent *in situ* hybridisation (FISH); depending on the technique used, landmarks can be localised to within a few kb. For example, Gläser et al. (1997) have used FISH mapping to order a number of Y-chromosomal short arm genes. However, it is very difficult to extend the range of these methods to above several hundred kb, and so complementary techniques must be used to obtain long-range maps.

Long-range physical maps may be assembled in a *top-down* manner, or *bottom-up*. Top-down refers to creating the map based primarily on the genomic DNA itself. This may be based on restriction enzyme sites of the genomic DNA. Bottom-up maps are aimed at reconstructing the order of landmarks on genomic DNA from previously cloned fragments of DNA, such as YACs (yeast artificial chromosomes; Burke, Carle & Olson, 1987), BACs (bacterial artificial chromosomes; for library construction and map assembly see Kim et al, 1996, and Asakara et al., 1997), plasmids, or cosmids. The fragments are ordered into contigs, to reveal the location of landmarks along the genomic DNA. One major assumption of this method is that the DNA structure is essentially unchanged between genomic and cloned forms. While this is largely true for the sequences of DNA (with the proviso that gross rearrangements can be fairly easily detected as chimaeric clones), any epigenetic effects, such as methylation patterns, are unlikely to be preserved; depending on the landmarks used, or the purpose for which the map is constructed, this may or may not be a problem.

Contigs are assembled on the basis that if two clones share a common landmark, then they overlap. Obviously, this assumes that the landmarks used are unique; however, in practice non-unique landmarks will be of use provided other data exists to localise the landmark as one particular copy. The types of landmark used include STSs (which do suffer from the uniqueness problem; particularly in some regions of the genome), genomic fragments used as hybridisation probes, subcloned probes derived from the ends of other clones (this being one of the main methods used for extending contigs where no previously-identified landmarks exist – known as ‘chromosome walking’;

see, for example, Kere et al., 1992), or restriction enzyme sites (of course, restriction sites are not unique, but a combination of enzymes will produce a unique pattern of sites on a clone).

The bottom-up method has the great advantage over the top-down technique of being based upon large clone libraries, which can be readily made available to other researchers in a standard, consistent form; this means that not only can many groups work together on assembling a map; but also once the map is assembled, the data will be directly relevant and useful to other groups, with no need to first relate the standard data to one's own idiosyncratic genome resource. One such widely available resource is the CEPH mega-YAC library; a collection of some 65,000 YAC clones, having a fourfold coverage of the human genome with an average insert size of 1 Mb (Cohen, Chumakov & Weissenbach, 1993)

However, there are disadvantages to assembling contig maps from cloned fragments. As mentioned above, there are the differences that may exist between the genomic and cloned versions of the same DNA sequences. Also, there are some regions of the genome which are unstable or even unclonable in the hosts used in library construction (Neil et al., 1990). To surmount this problem, it will be necessary anyway to resort to alternative libraries or the genomic DNA in order to close any gaps left in the contig map. Possibly a greater problem than unclonable sequences is the instability shown by certain sequences in the host cells – the DNA may initially be cloned intact, but over the generations some sequences (particularly repeated regions) are prone to be lost during cell division. This creates large problems in map assembly: not only are the sizes of such regions uncertain, but also a landmark may be judged present in a clone in initial rounds of screening, but lost during later propagation, thereby creating problems in the interpretation of results. A final cloning problem is the creation of chimaeric clones, bringing together sequences from separate genomic regions in the same clone. Care has to be taken to identify such sequences, and exclude them from the final map. Some cloning methods suffer less from these problems than others – certain strains of yeast are recombination-deficient, and so will not create so many chimaeric clones, or be so prone to instability as normal yeast strains (Le & Dobson, 1997). And methods such as BACs seem not to suffer so much from instability as do all yeast strains. However, the availability of these resources is

more limited than YAC libraries; and the insert size is smaller than the 1 Mb achieved with the CEPH YAC library.

Of course, the method chosen to construct a map depends on a number of factors: the resolution and detail desired, the landmarks to be used, the availability of resources, the current state of knowledge about the region under study, and the purpose to which the map is to be put.

Finally, the ultimate detail of mapping to be achieved by the HGMP is the entire base sequence of the human genome. This is projected to be completed by 2004, and will result in a vast corpus of information to be interpreted and analysed. This does not mean that lower-resolution maps are of no use, however: firstly, it is essential to assemble contig maps in order that clones may be chosen to sequence; secondly, despite the avowed intentions, the complete genome will not be sequenced, and so it is important to know where the main efforts should be directed, and which regions may be treated in less detail. In these regions which will not be sequenced, maps will be the only source of physical information. It may therefore be particularly important to have complete and accurate maps for these regions, perhaps more so than in the sequenced regions.

We shall now look at the mapping efforts as directed toward one specific human chromosome.

1.2 The human Y chromosome

The human Y chromosome is one of the smallest of the natural human chromosomes, being only some 60 Mb in size. For the bulk of the chromosome, genetic analysis is not possible (the exceptions being the small pseudoautosomal regions [PAR1 and PAR2] at the telomeres); thus, of necessity, the focus of Y chromosomal mapping is on the assembly of physical maps.

1.2.1 Y Chromosome evolution and function

The human Y chromosome is interesting to study for a number of reasons. Unfortunately, these same reasons also make the Y chromosome extremely difficult to

study. Firstly, it carries the sex-determining factor, *SRY*. Like other sex chromosomes, this leads over the course of evolution to a chromosome which carries very little else – presence of *SRY* selects against recombination of the sex chromosomes (aside from the pseudoautosomal regions, which on the human Y chromosome cover 2600 kb on Yp, and 320 kb on Yq, adjacent to the telomeres; see Rappold 1993), in order to avoid loss of the sex-determining factor. Following on from this, as the chromosome does not now recombine with a homologous partner, degeneration and loss of other coding regions will occur; and genetic functions of the proto-Y will migrate to other chromosomes as time goes on. This process would eventually lead to a Y chromosome devoid of all gene function aside from the *SRY* gene.

Working against this drive toward loss of genetic function is the tendency for genes involved in spermatogenesis and other male-specific processes to become localised to the Y chromosome. As the chromosome is, by definition, present and active in all male cells, this accumulation of male-specific genes is evolutionarily advantageous. According to the latest work (Vogt et al., 1998), genes and genetic functions found on the Y include *TSPY* (testis-specific protein on the Y), *DAZ* (deleted in azoospermia), *AZF* (azoospermia factor), and *GBY* (gonadoblastoma). Other research (Lahn & Page 1997) has detected a number of novel mRNA transcripts from the Y chromosome. It would be expected that many of these transcripts will play some role in spermatogenesis, while others appear to be X-Y homologous genes involved in housekeeping activities. Further research on these genes will reveal their functions.

Structurally too, the Y chromosome is different from other chromosomes. Much of the chromosome consists of repeated sequences – these include the pericentromeric satellite repeats, as well as most of the long arm heterochromatin, which is formed of a 5 bp satellite sequence (*DYZ1*), and a 2.4 kb repeat (*DYZ2*). Many rearrangements and duplications have taken place on the Y over the course of its evolutionary history, and large blocks of sequence can be found to be duplicated on the chromosome.

In addition to this, as the chromosome was originally the identical homologue of the X chromosome, substantial areas of the chromosome are regions of X-Y homology. The two pseudoautosomal regions are probably the only remnants of the ‘original’ Y chromosome, with the non-recombining homologous regions of the chromosome

being the result of numerous translocations from the X chromosome over the course of the evolution of the Y chromosome. There are at least 11 distinct X-Y homologous regions on the non-recombining part of the Y (Sargent et al., 1996). These represent various translocation events, the timing of which may be estimated by comparative studies of other primate Y chromosomes. For instance, the Yp / Xq21 block of homology represents the major block of homology on the short arm (Page et al., 1984), and is not found on the Y in chimpanzees or gorillas (Lambson et al., 1992), indicating that this is of relatively recent origin in humans. There are two blocks of homology in this region of the Y in humans, and molecular clock analysis of these Yp blocks (Schwartz et al., 1998) indicates that both blocks have diverged from the Xq21 region by the same amount. This suggests that the origin of these blocks was contemporaneous, and the block was subsequently split by rearrangements on the Y chromosome. Other lines of evidence lead to this conclusion as well – Sargent et al. (1996) found that the order of loci on these two blocks was highly conserved between the X and Y chromosomes, apart from an inversion of one group of loci.

The many other blocks of X-Y homology over the chromosome possess similar features, suggesting the history of the Y chromosome evolution is very complex, involving a large number of rearrangements and translocations. For example, Bickmore & Cooke (1987) studied the Yq11.1 / Xq26 region of homology, finding it not to be present on the Y chromosome of other primates. Careful analysis of these homologous regions can be used to study the history of the chromosome since the separation of the various types of primate, and to reconstruct the patterns of rearrangement on the Y chromosome.

The high level of repeated sequences, and similarities with the X chromosome, combine to make the Y chromosome very difficult to map effectively.

A final reason for interest in the Y chromosome is the uniparental inheritance pattern which it displays. In a similar manner to mitochondrial DNA, this inheritance can be used very simply to trace evolutionary relationships between populations – leading to a ‘Y chromosomal Adam’, (for one such study, see Zerjal et al., 1997) alongside the ‘mitochondrial Eve’. However, in order to make such studies and comparisons, it is necessary to have as large a number of polymorphic loci on the chromosome as

possible. The Y chromosome is relatively sparse in such polymorphisms as compared to the autosomes, probably due to the smaller number of Y chromosomes in the population (one quarter of the effective population size). Hence, detailed mapping of the chromosome is helpful to discover the best regions in which to search for novel polymorphisms to further refine evolutionary relationships uncovered via the Y chromosome.

The course of evolution of the Y chromosome in non-human mammals is being studied by a number of groups. Comparative studies have largely been carried out on primates and mice, revealing a number of proposed inversions and duplications presumed to have occurred in the course of human evolution. Also of interest are marsupial Y chromosomes, some of which are so small as to be nearly invisible cytogenetically. Study of these can lead to identification of essential Y-encoded genes, as well as those genes which have migrated to and from the Y chromosome over the course of its evolution.

1.2.2 Structural overview of the human Y chromosome

Physically, the Y chromosome may be divided into five regions: the PARs, Yp, the centromeric region, the Yq euchromatin, and the Yq heterochromatin. The salient features of each of these regions are given below, and summarised in figure 1.1.

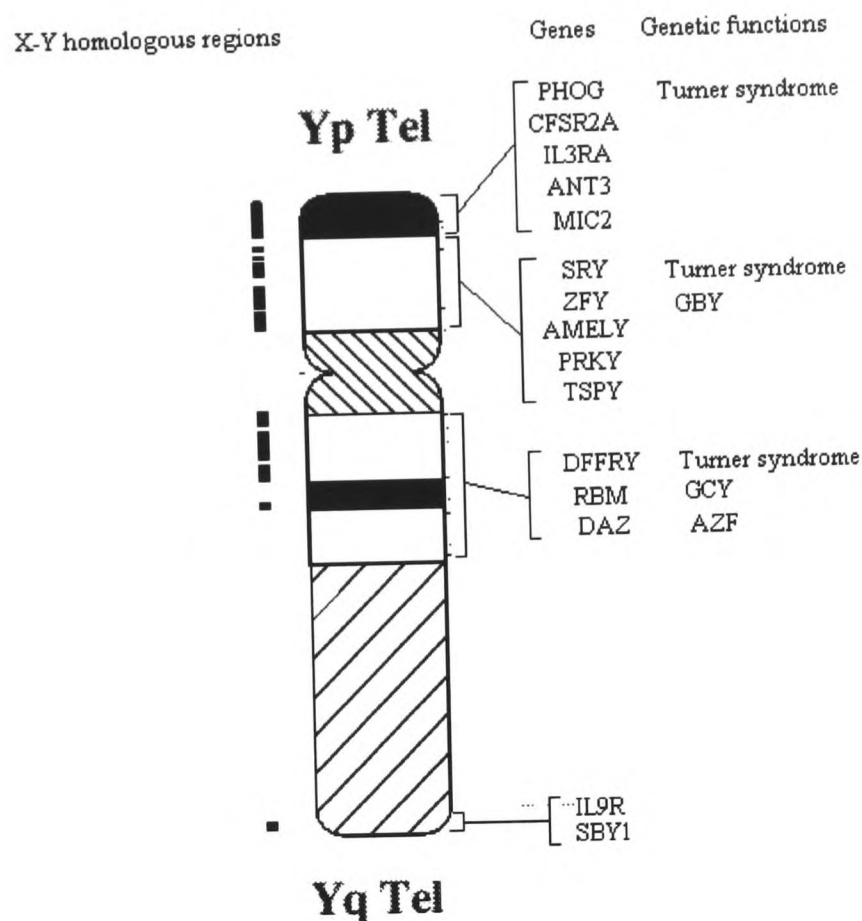


Figure 1.1: A summary map of the Y chromosome, showing the locations of selected genes and genetic functions referred to in the text, and marking the locations of X-Y homologous regions along the chromosome. (Based on Vogt et al., 1998)

PARs: As mentioned above, there are two pseudoautosomal regions on the Y chromosome. The Yp PAR covers 2.6 Mb, and has been mapped in some detail (Ried et al., 1995). This contains such genes as *CFSR2A* (a growth factor), *IL3RA* (the interleukin receptor), *ANT3* (adenine nucleotide translocase), and *MIC2*. For some time the existence of a genetic locus related to stature has been suspected in PAR1; recently two reports have found a gene, designated as *SHOX* or *PHOG* (Ellison et al., 1997; Rao et al., 1997); this gene may be responsible for some of the phenotypic features of Turner syndrome localised to various regions along the chromosome (see also paragraphs on Yp and Yq below). The *XG* gene has been identified as actually spanning the pseudoautosomal boundary, so transcription is terminated prematurely. It is unlikely that the Y chromosomal copy of this gene has a function, being merely the homologue of the functional X chromosome copy (Weller et al., 1995).

PAR2, the Yq PAR, has more recently been identified (Freije et al., 1992). This is a much smaller region, only 320 kb in size; and so far is known to contain two genes, *IL9R*, an interleukin receptor gene, and *SBY1*, a synaptobrevin-like gene.

Due to the nature of pseudoautosomal regions, they are amenable to genetic analysis, unlike the remainder of the Y chromosome. It has been found that 1 cM on the Y PARs is equivalent to a physical size of 53 kb (Henke, Fischer & Rappold, 1993); this reflects a recombination rate some ten times greater than the genome-wide average.

The Y chromosome short arm (Yp): Many genes have been localised to this region, in addition to several genetic functions (based on deletion mutations where no gene has yet been found). These include *SRY*, the sex-determining region of the Y. This is perhaps one of the most-intensively studied and hunted-for genes on the chromosome. *SRY* lies close to the pseudoautosomal boundary (PABY); the gene does not appear to be highly conserved between species, with the exception of a DNA-binding domain. It has been conclusively shown that it is this gene alone which is responsible for sex determination in mammals. Other genes in this region include *ZFY* (zinc finger Y), for a time thought to be the sex-determining gene, *RPS4Y*, an X-Y homologous ribosomal protein gene (Fisher et al., 1990), *AMELY* (amelogenin, Y copy; encoding a protein used in tooth development, with an X homologue, *AMELX*), *TSPY* (testis-specific protein on the Y), a repetitive sequence of 20-40 copies based on a 20 kb repeat unit, and found in two separate clusters, *TSPY1*, the larger, and *TSPY2*, a smaller cluster found adjacent to the alphoid repeats (see below), and *PRKY* (protein kinase Y; Klink et al., 1995), a homologue of *PRKX* on the X chromosome, which appears to also act as a hotspot for illegitimate recombination between the X and Y chromosomes. Genetic functions located to this region include GBY (gonadoblastoma), and also some effects of Turner syndrome (see Zinn, Page & Fisher, 1993).

The centromeric region: The major feature of this region is a block of alphoid satellite DNA; further discussion of the human centromere can be found in section 1.7 below.

Yq euchromatin: Several gene families or repetitive genes are found on the long arm of the chromosome. These include the *RBM* gene family (Ma et al., 1993), which has

around 30 copies on the Y, and appears to be involved in testis specific RNA splicing events. The *DAZ* gene family is a multicopy family, distinguished by varying numbers of the seventh exon, ranging from 7 – 24 (Saxena et al., 1996). The number of genes in the family is still unknown. As the name suggests (deleted in azoospermia), the gene may play some function in spermatogenesis. Non-repeated genes on the long arm include *DDFRY* (*Drosophila* fat-facets related gene on the Y; see Jones et al., 1996). Other genetic functions thought to be present in the region include further Turner syndrome effects, GCY (growth control on the Y) and additional AZF (azoospermia factor) loci, found in several separate locations (Vogt et al., 1995).

Yq heterochromatin: This covers some 20-30 Mb of DNA, and is composed of 5 bp repeats (Cooke, 1976), organised into 3.5 kb higher order repeats (Nakahori et al., 1986), and 2.4 kb repeats each containing an Alu sequence (Frommer, Prosser & Vincent, 1984). The reason for the presence of such a repeated region is unclear, when some Y chromosomes (notably the marsupials, and some humans) appear to be able to function without it. Other members of these repeated sequence families have pericentric locations, and it has been suggested that most of the long arm has undergone an inversion, to place previously pericentric sequences in a distal location.

Other features: Polymorphisms and homologies. As mentioned above, there are several regions of XY homology scattered about the Y chromosome. Most obviously, there are the PARs; as mentioned above, these are the remnants of the original homologue of the X chromosome. The remainder of the homologous regions represent numerous translocations from the X after the initial divergence of the two chromosomes. The blocks of homology are marked on figure 1.1; what this figure does not show is the fact that the equivalent regions on the X chromosome are scattered over the chromosome, rather than being in the same order on both chromosomes. This reflects the complex history of these regions, and the many rearrangements which have taken place on the Y chromosome after the transfer of sequences from the X. The main regions of homology are found on the Y short arm, where around 3 Mb of sequence proximal to the *ZFY* gene corresponds with the Xq21 region; along with a number of smaller regions of homology on this part of the Y – *ZFY*, *AMELY*, *PRKY*, and *GMGXY8* are all separate blocks of homology. In addition to these, the long arm has at least 6 blocks of XY homologous sequences, originating

from both arms of the X chromosome. The complex nature and organisation of these homologous regions is due to the severe reorganisation of the Y chromosome throughout its evolution.

A number of polymorphisms have been found on the Y, and are of great use in reconstructing phylogenies. There are several RFLPs (Casanova et al., 1985), a single Y chromosomal minisatellite (MSY1; see Jobling et al., 1994), and several microsatellites (Roewer et al., 1992). Other polymorphisms include divergence of the alphoid satellite repeat units, distinguishing at least 35 Y chromosome haplotypes (Santos et al., 1995). One of the largest studies has used 20 unique mutation events to classify 21 Y chromosomal haplotypes (Underhill et al., 1997).

1.2.3 Mapping the Y chromosome

In addition to the studies of specific genes, there have been a few large-scale physical maps focussing on the Y chromosome. One notable work has been that of Foote, Vollrath, Hilton & Page (1992) who have claimed to map the euchromatic region of the human Y chromosome.

1.2.4 The Foote et al. map

Foote et al. utilised STSs in order to map this chromosome. STSs are pairs of primer sequences which (ideally) define a unique PCR product. Thus, a rapid analysis of YACs can be performed by screening the YACs for the presence or absence of each unique PCR product. Two or more YACs sharing a particular STS must be derived from the same region of the chromosome, and therefore must overlap. In this way, a contig can be built up. The major advantage of STS screening is that it can be carried out in parallel, giving an assembled contig of YACs relatively rapidly.

Foote et al. constructed their own YAC library from a 49, XYYYYY cell line, and proceeded to screen this library by PCR for 160 different STSs. 234 YACs were isolated, with an average insert size of 580 kb. Further screening assembled 196 of these YACs into 10 separate contigs. The ends of the contigs were then used in order to generate further STSs for more screening; in this way, all 9 gaps were closed. Thus, the final result was a single YAC contig, covering the Yp region, the centromeric region, and the proximal end of Yq.

1.2.5 Problems with this map

Whilst this map is undeniably an achievement, and of utility in further work, the Foote et al. map has a number of flaws. The most important of these are discussed below.

Firstly, several of the STSs used in the screening process are not unique. For example, the loci sY60, sY61, and sY62 have been found to map to two separate positions on Yp. This will, at the very least, jeopardise the integrity and consistency of the map. While the total number of repeated STSs is unknown, there are at least these three, and possibly many more. The presence of repeated STSs can lead to gaps in the map not being noted, if two separate copies of the STS are treated as a single copy, and the intervening genomic sequences are not placed on the map; alternatively, if several repeats are interspersed with unique sequences, then the ordering of the unique sequences can be ambiguous or undetermined from the available data. Another problem with uniqueness was raised by the observation that the PCR assays used were poor at distinguishing regions of X-Y chromosome homology. Thus, some supposedly Y chromosomal YACs may be from the X chromosome, and contain non-homologous sequences which would obscure the map.

Comparison of the Foote map with the regions studied in detail in the current work is informative. On proximal Yq, the sequences 64a and KFC53 are known to be present in a repetitive region covering not more than 2 Mb in total. However, these sequences are not found on the Foote map, suggesting there is a relatively large gap in the map which has gone unnoticed. Proximal Yp, again, has problems. Foote et al. link the alphoid block to the 50f2/D (DYS7) and TSPY (sY60) sequences adjacent, as does the current work; and the *AMELY* block is linked to the large X-Y homologous region, again as has been done in the present work. One major flaw, however, is the linkage of the alphoid and *AMELY* blocks via the YACs yOX25 and yOX26. These two YACs are meant to have different STS contents, but according to our analysis appear to be identical (C. Tyler-Smith, pers. comm.); Foote kindly sent the earliest frozen stocks of the two YACs, which again are identical, both YACs corresponding to yOX25. This YAC contains the sY60 region, which would link it to the alphoid sY60 block; however, the YAC is actually from a third sY60 region, TSPY3. Thus the

alphoid and *AMELY* contigs are not actually linked by the Foote map, and there is a gap of unknown size between the two.

Other problems with the methodology of the mapping include the fact that not all of the Y-chromosomal YACs were assigned to the map. For instance, yOX81 is shown in this thesis to map to the *AMELY* contig, yet it does not appear on the Foote map. With respect to the detail of the mapping, although the STSs were grouped into subintervals by the team, within a subinterval the STSs were not ordered with respect to one another. This further reduces the utility of the map. Further, the library used was found to have a chimaerism frequency of around 59%. Chimaeras remain undetected in an STS-based approach, since the assay only detects specified regions of the chromosome. While this is an advantage in terms of the ease of constructing the map, and while it does not actually detract from the fidelity of the map itself, it is a practical disadvantage, since the cloned DNA available is not a faithful representation of the genomic DNA. Also, the map itself is not a faithful representation of the cloned DNA. Another practical disadvantage of this approach is that we do not know where the chimaeric regions lie on the YACs - thus, further characterisation of the YACs will be necessary for many future applications of the library. Furthermore, use of the YAC ends to derive further STSs may prove to be a problem if regions of chimaerism are not taken into account.

Historical contigs are also a problem with this map. These result when a YAC gives a positive result for an STS in the initial screening, only to have the region deleted during further rounds of growth. Thus, when the final screening is carried out, there may no longer be a YAC which represents the particular STS. So, a link between YACs is depicted, since it once existed, despite the fact that the link can no longer be found in the library.

The final problem with any STS-based map is that the use of STSs does not determine the sizes of anything. The size of the YACs is unknown (although this may be determined by means of PFGE); the spacing between the STSs is unknown, and hence the extent of overlap between YAC clones cannot be accurately determined; and also the length of the contig cannot be accurately gauged (although it can be estimated from the average YAC size, and the amount of redundancy in the library).

Thus, the Foote et al. map, which is one of only two major maps of the Y chromosome to date, has a number of shortcomings. Most importantly, it contains several gaps of unknown size on the regions of both the long and the short arms that we have evaluated in detail. Also, a number of the landmarks used are ambiguous, either in the sense of being non-unique, or in that the ordering has not been determined; and the map does not provide enough information as regards the make up of the clones used or some regions of the mapped DNA itself. The main flaw in the presentation of the map is that these defects were not recognised as being a problem. The second Y chromosome map (Jones et al., 1994), is also an STS-based map, and is largely based on information derived from Foote et al., and thus suffers from many of the same shortcomings, in addition to the absence of the actual centromeric region from this map.

As has been mentioned several times, the Y chromosome map lacks much of the centromeric region. In fact, to date no human chromosome contig maps have been produced which include the centromere, and link it to the flanking sequences on either side. The structure has, however, been very intensively studied, and the gross organisation is fairly well understood in a range of organisms. It seems appropriate at this point to include a digression on centromere structure and function in several organisms, including humans.

1.3 The Centromere / Kinetochore Complex

The centromere is defined as the *cis*-acting DNA element required for proper segregation of chromatids at mitosis and meiosis. Often this is interchangeably used with the term kinetochore; however, it is less confusing to reserve kinetochore to mean the assembled protein complexes which form around the centromeric DNA during cell division.

There are a number of distinct types of centromere / kinetochore complex, as regarded cytogenetically. The simplest is that of yeast, where no visible structure is present. Those of mammals are the 'classical' structure, with the primary constriction, a kinetochore disc, and a fibrous corona all present, as well as many microtubules seen to be emanating from the kinetochore region. Other types include the 'ball and cup'

kinetochore observed in plants and a number of insects; and the holocentric chromosomes of, for example, *Caenorhabditis*, where the entire surface of the chromosome serves as a centromere.

The mammalian centromere is visualised cytogenetically as the region of primary constriction on metaphase chromosomes. Under the electron microscope, the kinetochore can be seen as a trilaminar disc structure in mammalian cells. An idealised mammalian chromosome at metaphase is shown in figure 1.2.

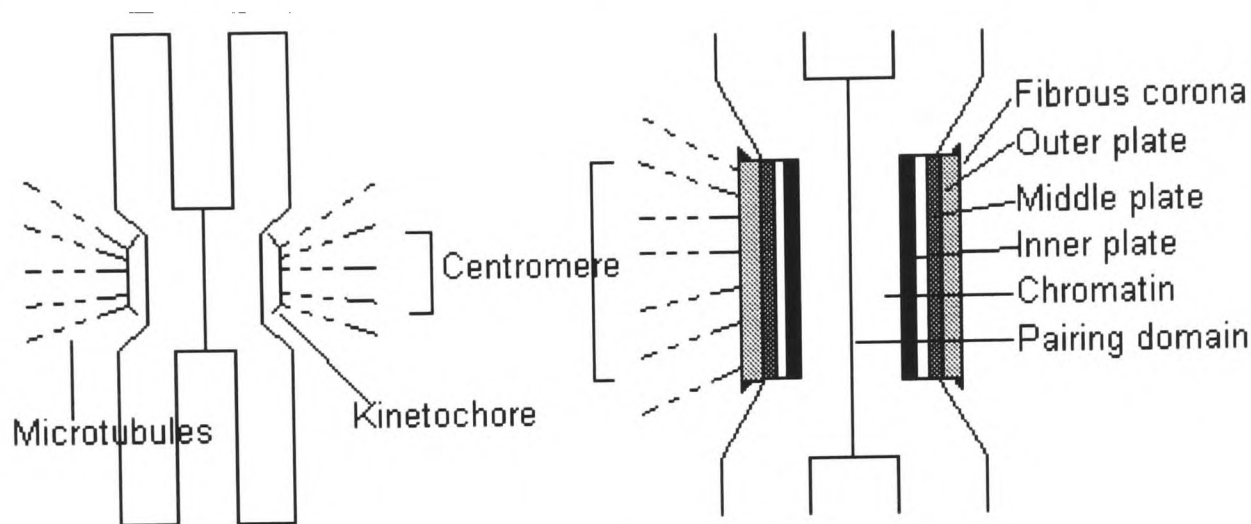


Figure 1.2: An idealised mammalian chromosome at metaphase.

The centromeres of a range of organisms have been studied; those of yeasts, *Drosophila*, and humans shall now be described.

1.4 Centromeres of *Saccharomyces cerevisiae*

1.4.1 Structural analysis

S. cerevisiae, or budding yeast, has had its centromere extensively characterised, due to its simplicity. The *S. cerevisiae* genome consists of 16 chromosomes, ranging from 260 kb to 3 Mb long. The functional centromere of these chromosomes has been found to be around 125 bp long, and contains short oligonucleotides that are highly conserved between different chromosomes (Clarke, 1990). The centromeric sequences were identified by means of an artificial chromosome assay. This relies on the fact that a plasmid containing an ARS (autonomously replicating sequence) will replicate in a yeast host; whilst in order to segregate efficiently at cell division, a centromeric sequence is necessary to stabilise the plasmid. Thus, a plasmid containing a selectable marker, an ARS, and a potential centromere sequence can be used to test candidate centromere sequences by transforming it into yeast.

It was found that a sequence some 125 bp long was sufficient to provide centromeric function. The deduced consensus sequence is shown in figure 1.3.

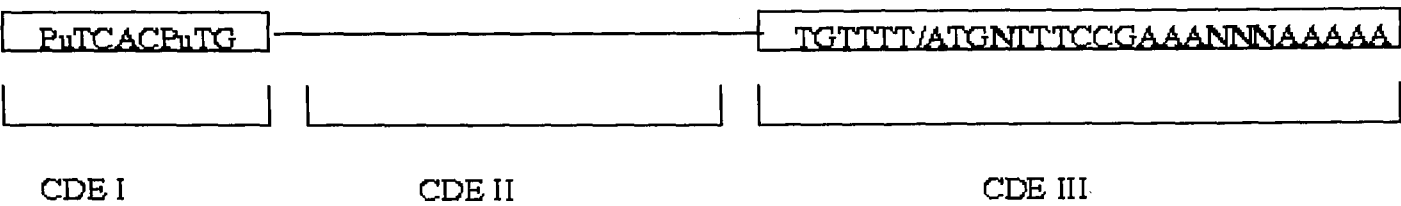


Figure 1.3: The consensus sequence of the *S. cerevisiae* centromere. Pu = purine; N = nucleotide.

The centromere consists of three distinct elements, termed CDE I, II, and III (for Centromere DNA Element). CDE II is an A-T rich (>90%) sequence of between 78 and 86 bases long, and is flanked by the consensus sequences CDE I and CDE III. The whole 125 bp sequence is itself contained in a 220-250 bp nuclease resistant region, which is surrounded by a number of nuclease hypersensitive sites.

1.4.2 Functional analysis

Altered centromeric sequences were tested by means of the artificial chromosome assay, or by replacing the wild-type centromere, in order to deduce the essential elements of the centromere. The main points are: the centromere is functional on any chromosome, rather than being chromosome-specific. Deletion of each of the CDEs results in the following: loss of CDE I results in a slight reduction of stability in mitosis, but random chromatid segregation during meiosis I; deletions and insertions in CDE II which alter the length of the sequence reduce segregation fidelity some 10-1000 fold, while a random 100% A-T sequence of the original length restores full centromeric function; hence it is the length and overall make-up of the CDE II sequence which is important, rather than the specific sequence information. CDE III is essential for centromeric function; in particular certain point mutations, including a change in any of the central CCG bases inactivate the centromere.

1.4.3 Proteins

The DNA sequences are necessary, but not sufficient, for a functioning centromere. The DNA interacts with a number of proteins in order to carry out chromosome segregation. Several such proteins have been identified in *S. cerevisiae*. CBF1

(Centromere Binding Factor) binds CDE I. It is a single polypeptide, of molecular weight 58 kDa (Cai & Davis, 1990). The precise function is unknown, but since CBF1 binding sites are found elsewhere in the genome, not necessarily at the centromeres, it is unlikely to be solely involved in centromere function. Mutations in the protein bear out this theory - mutants show poor chromosome segregation, as well as methionine auxotrophy.

The protein CBF3 is a 240 kDa complex composed of subunits A, B, and C (Lechner & Carbon, 1991). The complex binds to CDE III; and the mutations in CDE III which abolish centromere function also prevent binding of the CBF3 complex. Subunit CBF3-A has an ATP-binding ability, and the complex as a whole has molecular motor activity. This implies that the protein complex is involved in movement of chromosomes along microtubules during mitosis or meiosis. It has also been found that, *in vitro*, CBF3 will bind microtubules. However, pure CBF3 has very little microtubule-binding activity; and so it may well be that CBF3 requires additional (unknown) cellular cofactors in order to function efficiently. An additional subunit, SKP1, has recently been identified (Connelly & Hieter, 1996), which, unlike the other subunits, shows significant conservation between species. The human form of the SKP1 protein is part of the cyclin A – CDK2 complex, which is involved in cell cycle regulation, not centromeric function. However, mutation analysis of yeast gives SKP1 a role to play in both cell cycle regulation and centromere function.

CBF5 has also recently been discovered, through low-affinity binding to centromeric DNA *in vitro* (Jiang et al., 1993). The C-terminus of this protein bears similarities to the microtubule-binding domains of microtubule-associated proteins - and, again *in vitro*, the protein has been shown to bind to microtubules. In addition, temperature-sensitive underexpression mutants for CBF3 have their mutation suppressed by overexpression mutants of CBF5; suggesting that the two proteins have similar roles to play in mitosis.

The MIF2 protein (Meluh & Koshland, 1995), which has been shown to have homologies with the human centromere protein CENP-C (see section 1.7.3), has also been shown to interact with the CDE I sequence in yeast. Mutations of the *MIF2* gene

are observed to stabilise dicentric chromosomes, suggesting that disruption of MIF2 adversely affects chromosome segregation at mitosis.

Finally, the CSE4 protein (Stoler et al. 1995), which probably interacts with CDE II, has been shown to bear significant sequence homologies with both the histone protein H3, and the H3-like human protein CENP-A (see table 1.2, section 1.7.3). Furthermore, the H3-homologous region has been shown to be the minimal essential sequence of the protein to support cell viability.

These elements are the only ones that have been found to be involved in the centromeric activity of *S. cerevisiae*. This simplicity is one reason why the region has been extensively studied, and the activity of each part is well understood. However, this simple centromere is a very poor model for those of other eukaryotes. The DNA sequence equivalents of CDE I, II, and III have not been identified; homologues of the crucial proteins CBF3 A, B, and C have not been found, and the homologues of other CBF3-complex proteins that are known (e.g., SKP1) are not involved in centromere function. Thus the CDE III / CBF3 complex may best be regarded as a specialised *S. cerevisiae* structure, rather than a general model for eukaryotic centromeres – possibly evolved in order to allow permanent attachment of the centromere to the microtubules. This may be an evolutionarily derived feature, and microtubule capture the primitive condition.

1.5 Centromeres of *Schizosaccharomyces pombe*

The centromere of *S. pombe*, compared to that of *S. cerevisiae*, is much larger: the centromere of chromosome 2 covers some 80 kb of DNA. The fission yeast *S. pombe* also has much larger chromosomes - only 3 chromosomes, ranging from 3.5 to 5.7 Mb in size. An important difference between the two types of yeast is that, while the centromeres of *S. cerevisiae* remain attached to a single microtubule throughout the cell cycle, the centromeres of *S. pombe* have to undergo chromosome condensation and microtubule capture in order to undergo mitosis. The centromeres of *S. pombe* thus have a number of similarities with those of higher eukaryotes which *S. cerevisiae* lacks. The centromeres are variable; with each chromosome having a different centromeric sequence, much like the variation in centromeric regions found in humans. In addition, the centromere of chromosome 2, at least, can also be

polymorphic between strains of the yeast. Another similarity with higher eukaryotic centromeric regions is that those of *S. pombe* are composed largely of repeated sequences. As depicted in Figure 1.4, the centromere of one strain of chromosome 2 consists of four 14 kb repeated units, made up of subunits termed K (6.4 kb), L (4.5 kb), and B (<1 kb). Two of the 14 kb units are inverted around a central, 7 kb, unique sequence core. This core is, like CDE II in *S. cerevisiae*, A-T rich, and at least part is unique to each centromere. The central core of chromosome 2 is also immediately flanked by c.1.5 kb of inverted repeat, not found in the other centromeres (Clarke, 1990).

Centromere 1 is the smallest and simplest of the centromeres, being only some 40 kb in length, and bearing only a single set of inverted repeats; while centromere 3 can be much larger (up to 100 kb), and including as many as 12 sequential K repeated units.

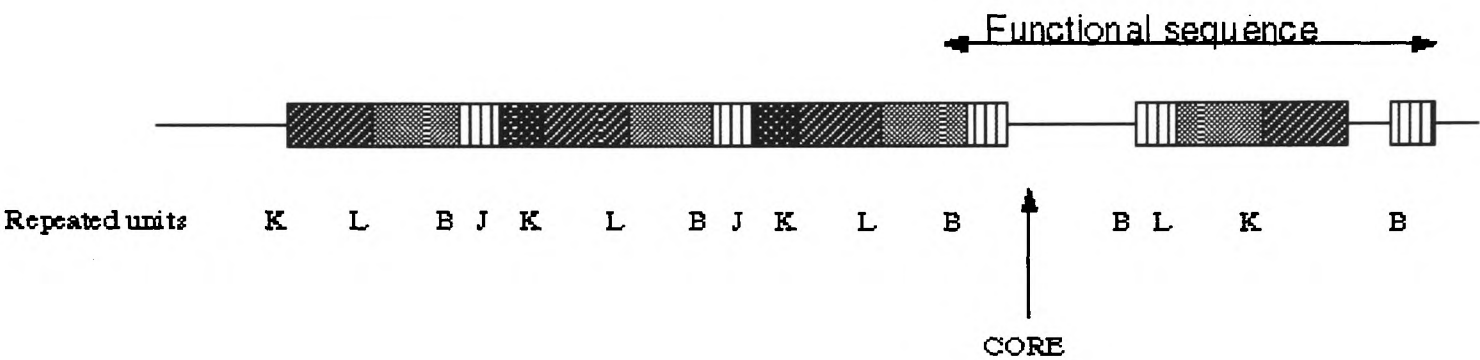


Figure 1.4: Centromeric region of the *S. pombe* chromosome 2.

The major similarities between all types of *S. pombe* centromeres are the inclusion of a central core; a version of the K repeat; and at least one inverted repeat flanking the core. The minimal functional sequence derived from centromere 2, as determined by an artificial chromosome assay, covers some 44 kb, and contains the 7 kb central core, the right hand B-L-K-B sequence, and around 3.5 kb of sequence on the left of the core. However, even this sequence does not have full centromeric function; full function has so far only been obtained when the entire region shown in figure 1.4 is used on the artificial chromosome.

One centromeric protein isolated from *S. pombe*, Abp1p (autonomously replicating sequence binding protein 1) has been shown to have homologies with the human protein CENP-B (Halverson et al., 1997), suggesting a similar function. When

overexpressed, the protein disrupts normal chromosome segregation, suggesting it plays some role in centromere function.

There have been recent suggestions that centromere function is in fact epigenetic, relying on modifications to the DNA being inherited, rather than the actual DNA sequence. In support of this theory, Ekwall et al. (1997), found that exposure of *S. pombe* to an inhibitor of histone deacetylase causes chromosome loss, as the acetylation of histones H3 and H4 is raised. This increased level of acetylation is, interestingly, inherited by subsequent generations of cells, even after the inhibitor of the deacetylase has been removed. These daughter cells also experience a greater rate of chromosome loss. This experiment suggests that the acetylation state of these histones is an epigenetic marker for centromere maintenance, rather than being purely affected by the DNA sequence.

Why is the centromere of *S. pombe* so much larger than that of *S. cerevisiae*? The answer to this question is unknown. Certainly, the chromosomes of *S. pombe* are larger than those of *S. cerevisiae*; this may necessitate the capture of more microtubules per centromere (the centromere of *S. cerevisiae* only binds a single microtubule); thus, the centromere may have to be larger. Alternatively, it is known that the centromeres of *S. cerevisiae* remain attached to the microtubules for the majority of the cell cycle; thus, since there is not such a need for the centromeres to capture the microtubules in this yeast, they may be smaller without sacrificing efficiency. However, this is still speculation.

1.6 Centromeres of *Drosophila melanogaster*

The makeup of *Drosophila* centromeres, and the localisation of essential sequences, has only recently been elucidated. Le, Duricka, & Karpen (1995) studied the organisation and sequence composition of the centromeric region of a non-essential *Drosophila* minichromosome, derived by inversion and deletion from the X chromosome. This chromosome, Dp 1187, is some 1.3 Mb in length, and so is amenable to analysis by PFGE to a much greater extent than natural chromosomes.

Restriction mapping and probing of a range of radiation-induced mutations revealed an interesting picture - the centromeric region of Dp 1187 consisted of 'islands' of complex DNA amongst a 'sea' of simple satellite sequences (see figure 1.5). The

islands were given the names Tahiti, Moorea, and Bora Bora. The island Bora Bora may, in fact, be two separate islands of 40 kb and 100 kb, separated by a small region of satellite sequence; this was impossible to determine from the data obtained.

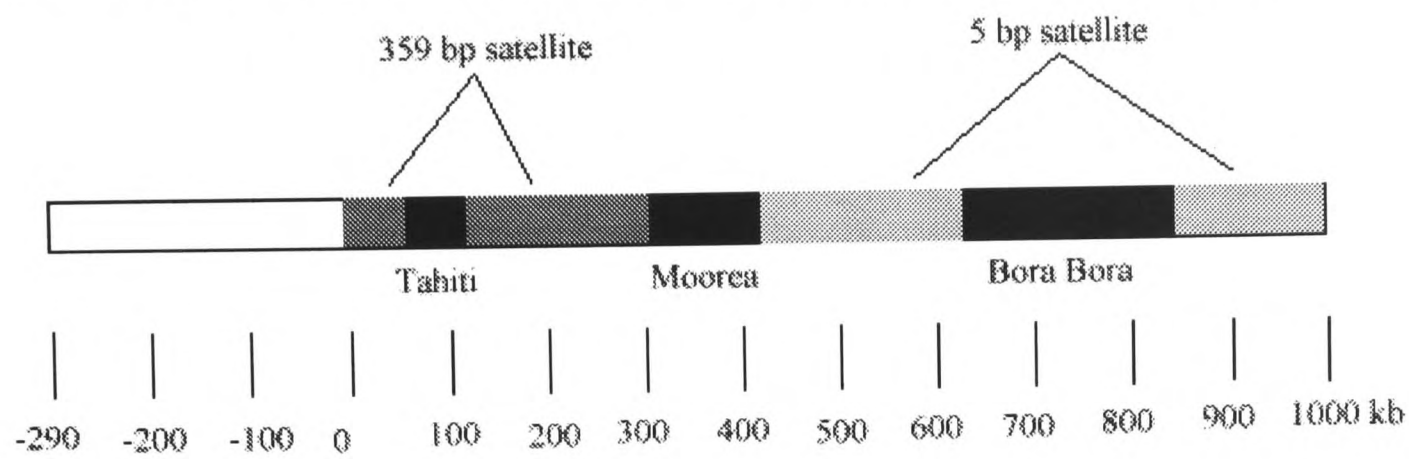


Figure 1.5: The centromeric organisation of the *Drosophila* Dp 1187 minichromosome.

As Dp 1187 is a minichromosome, it was thought that the rearrangements which led to its formation may have distorted the centromeric organisation. For this reason, the study then examined natural chromosomes, in order to determine if they, too, had the island / sea organization. All three islands were found to be present in natural X chromosomes, although Tahiti and Moorea were inverted with respect to one another. Sequencing of the Tahiti island determined that it shared sequence homologies with the *Doc* transposon; hence, it is possible that the complex DNA sequences of *Drosophila* centromeres are the result of transposition events.

Once the sequence organization of the centromere was determined, it was natural to investigate which sequences are essential for centromeric function. Murphy & Karpen (1995) studied the same minichromosome, generating a panel of deletions for functional analysis. It was discovered that the region of Bora Bora, from position +580 to +800, was present in all functioning chromosomes, and absent from unstable chromosomes. However, in order to obtain chromosome transmission of the order of natural chromosomes, a further 200 kb of flanking satellite sequence was required from either side of the Bora Bora island.

Further study of the organisation of the centromere (Sun, Wahlstrom & Karpen, 1997) has recently shown that the 420 kb essential region consists largely of satellite sequences interspersed with short stretches of more complex DNA. These complex

regions appear to be a number of complete, intact transposons. However, the transposons are not unique to centromeric regions, and nor are the satellite sequences. Neither is any sequence shared by all centromeric regions. This offers support for an epigenetic model of centromere function, as has already been proposed for yeast (see above). Confirmation of this theory, however, awaits experimental evidence.

1.7 Centromeres of *Homo sapiens*

1.7.1 Structural analysis

Despite having been intensively studied, far less is known about human centromeres than those of the yeasts discussed above. Humans have 46 chromosomes, of a size ranging from 60 Mb to 150 Mb. The functional centromere of a human chromosome is hard to define precisely, as no conclusive functional assay exists. Cytogenetically, the centromere is the site of the primary constriction; or, taking the size of the major alphoid array as the size of the centromere, this spans some 1 to 3 Mb, far larger than those of the yeasts. Like *Schizosaccharomyces pombe*, the centromeric regions of human chromosomes consist mainly of repeated sequences. These are mainly satellite DNA, of which a number of families have been identified; these are summarised in table 1.1, below (adapted from Tyler-Smith & Willard, 1993).

Table 1.1: Major satellite families of human centromeric regions		
Repeat unit	Name	Location
5 bp	Satellites 2 and 3	Most (or all) chromosomes
42 bp	Satellite 1	Chromosomes 3, 4, 13, 14, 15, 21, 22
48 bp		Chromosomes 13, 14, 15, 21, 22, Y
68 bp	β satellite	1, 3, 9, 13, 14, 15, 21, 22, Y
171 bp	α satellite (alphoid)	All chromosomes

Each satellite family also consists of a number of subfamilies. These variants are often chromosome-specific, or common to a subset of the chromosomes. In addition to these satellite sequences, the bulk of the remainder of the centromeric region is thought to consist of several interspersed localised repeated sequences. There is no evidence so far for the existence of any unique sequences in the centromeric region of human chromosomes.

Several of the human centromeric regions have been analysed; a long-range map of the Y chromosome centromere, depicting the location and type of the satellite sequences, is shown in figure 1.6 (Adapted from Tyler-Smith & Willard, 1993).

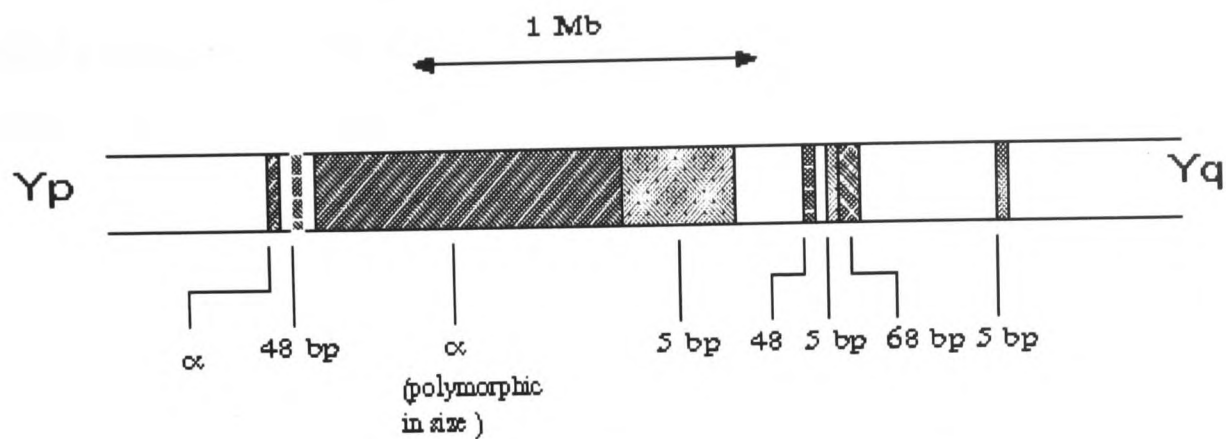


Figure 1.6: Satellite sequences of the human Y chromosome centromeric region.

The size of the Y chromosome centromeric alphoid array is polymorphic, ranging from 130 kb to 1.7 Mb. However, the general arrangement of sequences is similar in the other chromosomes. Cytogenetically, the alphoid DNA is to be found at the primary constriction of all natural human centromeres, as localised by FISH.

1.7.2 Functional analysis

Unlike the yeasts, no simple artificial chromosome assay is available for use in mammalian cells. As a result of this difficulty, the functional sequences required for human centromeric ability have still not been definitively identified. However, there is strong support for the alphoid DNA sequences being an essential part of the centromere. This support comes from a number of lines of evidence, foremost of which is the presence of alphoid DNA at all human centromeres; the only sequence to have been shown to have this distribution.

In addition, a number of transfection experiments have been carried out. Haaf, Warburton & Willard (1992) transfected African Green Monkey cells with alphoid DNA derived from human chromosome 17. Of those cells in which several Mb of human DNA successfully integrated into the host genome, a high proportion of dicentric chromosomes were observed, suggesting that the presence of a large block of alphoid DNA is enough to disrupt segregation, suggestive of some centromeric activity. A cell line in which only some 50 kb of human DNA had integrated did not display any abnormalities. Caution must be exercised, however, as it was still possible that some other, as yet unidentified, monkey sequence may have integrated along with the alphoid sequence, and this may instead be responsible for centromere formation. Stronger evidence was provided by the work of Harrington et al. (1997), who

transfected a human cell line with synthetic alphoid DNA arrays to create stable novel centromeres; this work is described more fully in section 1.8.1 below.

However, it is entirely possible that, while alphoid DNA may play some role in centromere function, it may also be necessary to have some other sequences present. A number of studies lead to this conclusion. Analysis of naturally occurring deleted Y chromosomes (Tyler-Smith et al., 1993), showed that all the deleted chromosomes which still segregated efficiently had a block of DNA in common. This block of DNA, which contains the centromere, was found to consist of a few hundred kilobases of alphoid DNA plus around 300 kb of adjacent short arm sequences. This centromeric region can be as small as c.500 kb; although the alphoid DNA is polymorphic. However, this study suffers from the disadvantage that only natural breakpoints were analysed - thus, the localisation of centromeric function may not be as precise as could be hoped.

This problem was avoided in the work of Brown et al. (1994), who used telomeric DNA in order to fragment the Y chromosome at a breakpoint in the interval defined by Tyler-Smith. This resulted in the production of two deleted chromosomes, one of 140 kb of alphoid DNA plus the Yp arm; the other consisting of 480 kb of alphoid DNA plus the Yq arm. Both chromosomes were found to segregate efficiently; however, the Yq derivative was observed to lag at anaphase in a proportion of dividing cells. Thus, since the only sequence shared by the two chromosomes was alphoid DNA, this sequence must be sufficient for segregation of chromosomes. But, as the Yq derivative behaved aberrantly at anaphase, some additional sequences on Yp must also be necessary for complete centromeric function.

This suggestion is also supported by the work of Larin et al. (1994), who integrated alphoid-containing YACs into human cells. The integrated alphoid DNA caused the production of a cytogenetic constriction, in addition to segregation abnormalities, suggesting the production of a centromere. However, the lack of actual chromosome breakage at mitosis suggested that the integrated DNA did not produce a fully functional second centromere; presumably, some further DNA sequences are required to obtain full centromeric function.

Doubt is cast on the essential role of alphoid DNA by studies of naturally occurring human mutant chromosomes. Most such rearrangements still carry alphoid DNA at the active centromere, supporting the claims made for the sequence; however, a number of chromosomes have been found to segregate normally, and yet carry no detectable alphoid sequences (Vaillaire et al., 1993; Depinet et al., 1997). A study of a chromosome 10-derived marker chromosome (du Sart et al., 1997) again found no alphoid DNA to be involved in the neo-centromere; but also determined there was no sequence difference between the neo-centromeric region and the region of chromosome 10 from which it was derived. They postulate an epigenetic effect such as the initial assembly of the DNA – protein kinetochore complex is responsible for the activation and maintenance of the centromere. Epigenetic effects have been demonstrated in centromere-containing plasmids of *S. pombe* (Steiner & Clarke, 1994; see also section 1.5 above for an epigenetic effect in native *S. pombe* centromeres), whereby inactive centromeres are heritably converted to active centromeres. The mechanism proposed by Steiner and Clarke is that the assembly of the DNA – protein complex is brought about by specific folding of the centromeric DNA sequences. It seems entirely possible that such a mechanism may act in human chromosomes as well.

What is certain about the centromeric sequence, however, is that it is very unlikely to be a unique sequence. A rearranged chromosome 17 was found to be split into two parts, both of which had a functioning centromere (Andersen, Tommerup & Koch, 1990). Thus, as the centromeric sequence can be split, both parts retaining centromeric function, the necessary sequence is most likely to be a block of repeated sequence.

1.7.3 Centromeric proteins

Several proteins have been identified in human cells which interact with the centromeric region of chromosomes. These were initially identified by means of antisera derived from patients with the autoimmune disease CREST (Calcinosis, Raynaud's phenomenon, oesophageal dysmotility, sclerodactyly, telangiectasia). A number of proteins were isolated, termed CENPs (for CENtromeric Protein). Not all of these have yet been extensively analysed; a brief summary of what is known about

some of the CENPs is given in table 1.2, and the discussion below concentrates on the best-characterised DNA binding proteins.

The protein CENP-B was the first to be studied in detail. It was found to bind to a 17 bp motif found in a few monomers of most alphoid sequences, termed the CENP-B box (Masumoto et al., 1989). Since alphoid DNA is found at all human centromeres, this protein was thought to be essential for assembly of the kinetochore complex. However, as yet, it seems that the alphoid DNA found on the human Y chromosome (as well as the equivalent satellite of many other mammalian Y chromosomes) lacks this CENP-B box motif. Furthermore, use of CENP-B antibodies has failed to detect CENP-B at a number of rearranged centromeres (in addition to the natural Y chromosome centromere); whilst CENP-B has been detected at several inactive centromeres of dicentric chromosomes. Thus it seems to be neither necessary nor sufficient for centromere function. Possibly it has a function in heterochromatin formation, or its relationship to transposases allows it to produce nicks adjacent to CENP-B boxes and promote the maintenance of alphoid DNA in the genome (Kipling & Warburton, 1997).

CENP-A, as a homologue of histone H3, seems to be involved in packaging of the centromeric DNA, and may well specifically mark the centromere, distinguishing it from the remainder of the chromosomal DNA, packaged with standard histone proteins. The work of Vafa & Sullivan (1997) shows CENP-A to be associated with α -satellite DNA sequences; however, the CENP-A does not simply recognise a sequence from the α -satellite DNA, as CENP-A has also been shown to localise to the centromeres of non-alphoid-containing neo-centromeres (Warburton et al., 1997). This work also specifically located CENP-A on the inner kinetochore plate; and also to only active centromeres, even when the inactive centromere contained α -satellite sequences. Again, it seems likely that some sort of epigenetic effect regulates the assembly of CENP-A to active centromeres, whatever their sequence composition. CENP-A itself could even be the epigenetic marker, as it is known to remain associated with DNA in bovine sperm (Palmer et al., 1991), whilst other histones are removed during gametogenesis.

CENP-C shows some homologies with MIF2, a yeast protein which binds microtubules and is essential in centromere function. The human protein may bind DNA, but no specific target sequence has been identified. Fukagawa & Brown (1997) used conditional knockouts of CENP-C in chicken cells to study the function of the protein, and found that CENP-C is necessary for chromosome movement in anaphase, but is unnecessary in other parts of the cell cycle; making it likely that the protein has much the same role as its yeast homologue. However, antibody microinjection experiments (Tomkiel et al., 1994) localise the protein in human cells to the inner kinetochore plate, and suggest that function of the protein may be necessary to maintain the organisation of the kinetochore complex.

Table 1.2: Selected human centromeric proteins			
Protein	Size	Comments	Reference
CENP-A	17 kDa	Similar to histone H3	Palmer et al., 1987
CENP-B	c. 80 kDa	Binds to CENP-B box of alphoid DNA as a dimer	Masumoto et al., 1989
CENP-C	140 kDa	Possibly involved in kinetochore assembly and maintenance. Molecular motor function? Present at same levels in all centromeres	Sugimoto et al., 1994
CENP-D	50 kDa	Homologous to RCC1 – regulator of chromosome condensation	Bischoff et al., 1990
CENP-E	250 kDa	Homologies with kinesin, a molecular motor	Yen et al., 1992
INCENPs (INner CENPs)	133, 155 kDa	Localised to the pairing domain between chromatids	Cooke, Heck & Earnshaw, 1987
PJα	10 – 15 kDa	Binds to a subset of alphoid monomers; function unknown	Gaff et al., 1994

1.8 Analysis of centromeres

1.8.1 Why study centromeres?

The reasons for desiring to analyse the centromeric region of humans are manifold. Once the structural analysis has been carried out, then it will be possible to isolate and identify potential candidate sequences for centromere function. As has been discussed, while alphoid DNA appears to be necessary for some centromeric function, it is by no means sufficient for accurate chromosome segregation. Analysis and definitive study of the make-up of the functional region should be of great assistance in the attempt to construct a functioning mammalian artificial chromosome (MAC). Like the YAC technology which has preceded it, the ability to make a MAC would bring about great advances in genetic research. Not only would we have a system in which to maintain cloned copies of DNA sequences in mammalian cells (potentially

much larger even than YACs); but it would be possible to study the interaction of these sequences with proteins expressed *in vivo* in the mammalian cell; as well as the expression patterns of these artificial genes. Furthermore, it is to be hoped that such a system would lead to great advances in the realms of gene therapy, by making it possible to introduce a functioning copy of an inactive gene into a patient's cells, with great accuracy and reliability. For this reason, too, the centromere is an important region to analyse.

Several disease phenotypes have been shown to map to the centromeric interval of the Y chromosome (for example, azoospermia); thus, genes must be present in this region. The centromere itself, however, is heterochromatic, a feature which is known to prevent gene expression. Thus, it will be of great interest to localise potential genes in the region, and determine how close to the centromere itself genes can reside and still be expressed - there is little point constructing a MAC for gene therapy with an inactive gene.

Harrington et al. (1997) claim to have produced a basic human artificial chromosome, achieved by transfecting cell lines with DNA fragments including synthetic telomeres, artificial alphoid DNA arrays, and genomic DNA; in some of the cell lines, these fragments combined to form stably replicated artificial chromosomes. However, due to the method used to assemble these chromosomes, very little can be said with any confidence about the essential factors in making artificial chromosomes. Also, as the sequences are assembled more or less randomly by the host cell itself, there would appear to be no possibilities of control or manipulation of the MACs, or of incorporating desired genes with any accuracy. A further disadvantage of this system is that the introduction of free telomeres into the cell leads to recombination and truncation of native chromosomes with a high frequency – an undesirable effect for a MAC to have.

An alternative approach to creating MACs is shown by the work of Heller et al. (1996), who used this truncation of native chromosomes by targeting free telomeres to the human Y chromosome, and thereby fragmenting it. In this way, they created a series of minichromosomes, down to 3.5 Mb in size, which could be used as the basis of a MAC. Again, however, this approach does not reveal anything further about the

necessary centromeric sequences, relying as it does on whichever sequences are already present on the native Y chromosome.

Detailed study of the naturally occurring centromeres is essential if we wish to understand the essential sequences for centromere function, and thereby hope to create MACs with more precision and functionality than those which exist at present.

In addition to these reasons, the Human Genome Mapping Project is preparing for a large-scale undertaking to sequence the entire genome. In order to be in a position to carry out this research, it is essential to construct sequence-ready large-scale maps of cloned DNA. At present, no sequence-ready map of a centromeric region has been constructed.

Over the last year or two, our view of the human centromere has changed substantially. We no longer think that, if a certain DNA sequence is present, then a centromere will form. Instead, we think of the centromere as a DNA – protein complex that perpetuates itself epigenetically, but can occasionally become inactive (losing essential protein components), or move to another position on the DNA (creating a neo-centromere). There are therefore new questions in the field: is there anything special about centromeric DNA? Neo-centromeres are often less efficient than natural centromeres, so it seems likely that natural centromeric DNA does have special features, and we therefore want a description of centromeric DNA structure. If centromeres can move, especially if they can creep along the DNA, why do they normally remain in a fixed position? Is there some feature of the DNA which prevents movement; or is there some protein involved in preventing movement? If there is a protein present, how does it attach itself to the centromere – DNA or protein interaction? Rather than just raw sequence analysis of the DNA to discover ‘the unique centromeric sequence’, these are the questions which must be addressed in future centromere research.

1.8.2 Problems with centromeric analysis

The bulk of large-scale human genome analysis and cloning has been carried out by means of the YAC system (Burke, Carle & Olson, 1987). Whilst this is sufficient for many applications, it encounters problems when applied to centromeric DNA regions.

Prime amongst these is the absence or instability of certain classes of DNA sequences in the yeast host (Neil et al., 1990). Such classes include repetitive DNA, which is prone to recombination in the yeast host. Since the bulk of the human centromeric region consists of repetitive sequences, this may prove to be a major problem. In order to circumvent this problem, recombination-deficient host strains may be used; however, at present none of the widely available YAC libraries have been produced in such a yeast strain.

The main method of circumventing this problem must be to use genomic DNA for the analysis. Being obtained directly from the genome of the organism being studied, it should not have undergone any rearrangements or alteration which would obscure the *in vivo* structure of the DNA. Possible methods of analysing genomic DNA include restriction mapping of whole cell DNA (although this in itself has difficulties for interpretation and analysis of results); or visual mapping by means of FISH techniques, in order to localise probes directly in the genome. Unfortunately, conventional FISH on metaphase chromosomes has a low resolution, and so techniques for high-resolution FISH must be used.

These problems relating to the difficulties of constructing maps of centromeric regions also apply to the repetitive regions elsewhere on chromosomes, such as make up a large part of the Y chromosome.

1.9 Purpose of this project

As section 1.2.5 showed, the mapping of the Y chromosome to date is not sufficient for several reasons. Thus, there is the need for further mapping work to be carried out. This laboratory has begun work on constructing a map of the pericentromeric 10 Mb of the Y chromosome. At the start of this project, the map consisted of 6 separate contigs, constructed from the readily available clones, spanning 10 Mb. The work has been carried out with YACs from the CEPH, OXEN, and St. Louis libraries. Some regions of the chromosome have been mapped in detail: around 200 kb of pericentric Yq (Cooper, Fisher & Tyler-Smith, 1992), and 500 kb of pericentric Yp, including alphoid DNA-containing YACs (Cooper, Fisher & Tyler-Smith, 1993).

There are three possible explanations for the presence of gaps in the map. One is that the gaps are covered by clones in the YAC library, but they have not yet been reached by chromosome walking, due to an absence of probes from the region at present. It is thought that the gaps between the TSPY, *AMELY*, and alphoid contigs; and the 14A3C and GMGY6 contigs (see figure 1.7) are of this nature. Such gaps are expected to yield to closure without much difficulty. The second type of gap is due to the complete absence of any clones from the region in the library. The 68 bp contig is thought to be separated from the other contigs by a gap of this sort. The clones may not be present in the library for a number of reasons - instability of certain repeated sequences in YAC hosts; or the effective unclonability of other sequences may explain these gaps. Such gaps will prove more difficult to close, although it is expected that using such techniques as detailed below will enable all gaps to be closed. A third explanation is that YACs from the region are present in the library, but do not contain Y-specific sequences, only repeated sequences. This is a possible explanation of the 68 bp – 14A3C contig gap.

The current map uses restriction enzyme digestion as the mapping method, rather than STSs. While this is more labour-intensive than PCR-based techniques, and has a much lower throughput, it does have many advantages over other techniques. Each YAC mapped confirms the evidence of previous YACs. The sizes of the YAC itself and the overlap between clones are directly determined by the mapping method. Chimaerism is more obviously shown; as well as the parts of the YAC which are chimaeric. The YACs are automatically oriented and located on the map, in terms of physical location - this gives a good starting point for 'walking' in order to extend the contigs. And the map itself is a direct representation of what is available in the cloned DNA; the clones of which are immediately available and accessible for further study. Furthermore, this map is entirely independent of the STS-based maps, and so can be used either to confirm the information thus obtained, or to highlight any discrepancies which may arise.

A summary of the map before commencing this study is shown in Figure 1.7.

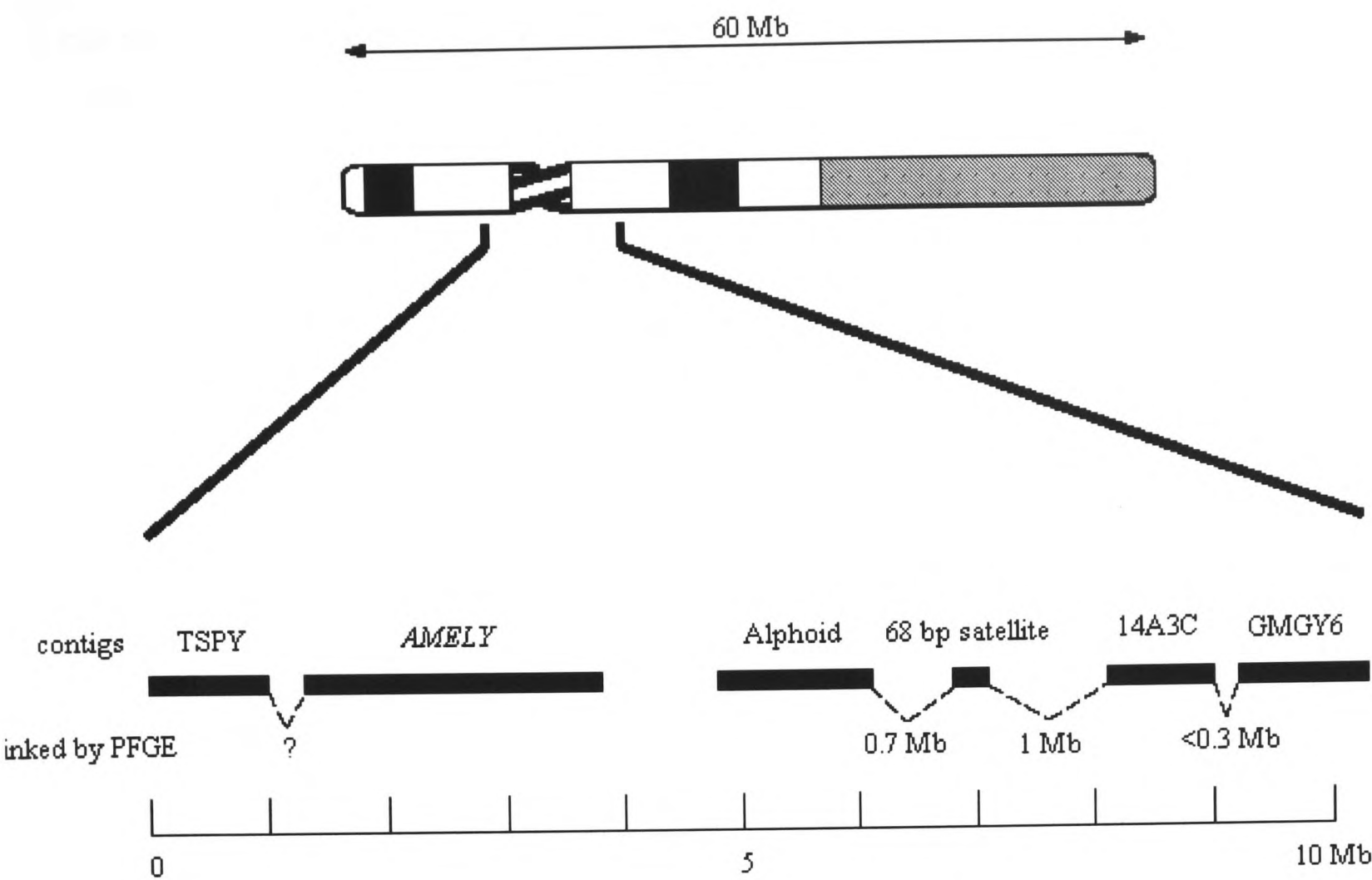


Figure 1.7: Summary of the YAC contigs at the outset of this work.

It is the aim of this current project to complete the mapping work which has been carried out on the centromeric region of the human Y chromosome. Ideally, the separate six contigs of the map from this laboratory will be united into a single contig, spanning the entire 10 Mb. This map will then be available as a resource on which to base further work. The bulk of this mapping project will be carried out using the publicly available YAC libraries. After screening the whole library with several probes from the region of interest, YACs will be identified from this region, and subjected to further analysis in order to characterise them and link the separate contigs once they are placed on the map.

After one round of YAC isolation and characterisation, the information gained will be used to return to the initial library and isolate further clones of interest.

Thus, the major aim of this project is to complete a physical map of the human Y chromosome pericentromeric region. In addition, it is hoped to study other aspects of the chromosome: in addition to the physical map, known genes will be localised on

our map, so relating our work to that carried out by others in the field. Abnormal Y chromosomes may be studied, combined with the gene localisations, to try to determine functions of the genes, and where any further as yet unidentified genes may lie. Finally, it is hoped that polymorphisms may come to light during this study, to assist in the efforts to reconstruct the phylogeny of the human Y chromosome.

The structure of this report is as follows: Chapter 2 describes the materials and methods used in carrying out the experiments of the work. Chapters 3 and 4 present the raw restriction mapping data of, respectively, the long arm and the short arm of the chromosome. The following chapters discuss the more detailed analysis of specific regions, focussing on localisation of some of the Y chromosomal genes, detection of polymorphisms, and some studies undertaken on aberrant Y chromosomes, to try to identify the events leading up to certain mutations in the phenotype. Finally, the thesis concludes with a discussion of the results obtained, and their relevance.

Chapter 2: Materials and Methods

2.1 Growth and storage of bacterial, yeast, and mammalian cells

2.1.1 Bacterial Cells

2.1.1.a Media and reagents

YT Broth: 16 g/l tryptone
 10 g/l yeast extract
 10 g/l NaCl

2.1.1.b Growth and storage

Cells were grown in liquid YT broth (+ 100 µg/ml ampicillin, or 30 µg/ml kanamycin where required) at 37°C on a shaker table. To store, aliquots were mixed with glycerol, to 50% (v/v), and frozen at -70°C. All cell lines were *E. coli* XL1-Blue cells.

2.1.2 Yeast Cells

2.1.2.a Media and reagents

SD Medium: 6.7g/l yeast nitrogen base
 20 g/l glucose
 55 mg/l adenine
 55 mg/l tyrosine
 70 ml/l 20% casamino acids
SD Agar: SD medium + 10 g/l agar

2.1.2.b Growth and storage

Colonies were grown on SD agar, then picked out into SD medium, and grown at 30°C on a shaker table. Agar plates were sealed with Nescofilm and stored at 4°C for up to 6 months. For longer-term storage, aliquots of SD medium + cells were mixed with glycerol to 50% (v/v) and frozen at -70°C. All yeast cells grown were AB1380 cells.

2.1.3 Mammalian Cells

2.1.3.a Media and reagents

RPMI 1640 medium (Gibco BRL) +10% foetal calf serum (Globepharm), + 20 mM glutamine, + 100 µg/ml kanamycin

2.1.3.b Growth and storage

Human lymphoblastoid cells were grown in RPMI 1640 medium at 37°C in a humidified atmosphere containing 5% CO₂. Cells were stored in liquid nitrogen by pelleting the cell suspension, and resuspending in freezing medium (95% foetal calf serum, 5% DMSO). Cells were aliquoted into 1 ml tubes, and placed at -70°C overnight, before being stored in liquid nitrogen.

2.2 Probes used in this work

A list of probes used in this work, and their sources, is given in table 2.1, below.

Table 2.1: List and source of probes used in this work		
Probe	Source	Reference
14A3C	C. Tyler-Smith, Oxford	Fisher, 1992
109e	C. Tyler-Smith, Oxford	Unpublished
116e	C. Tyler-Smith, Oxford	Unpublished
219e	C. Tyler-Smith, Oxford	Unpublished
RBF5	C. Tyler-Smith, Oxford	Fisher, 1992
GMGY6	C. Tyler-Smith, Oxford	Affara et al., 1986a
92e	C. Tyler-Smith, Oxford	Unpublished
64a	J. Weissenbach, Paris	Guellaen et al., 1984
KFC52	C. Tyler-Smith, Oxford	Cooper, 1992
68 bp satellite	C. Tyler-Smith, Oxford	Cooper, 1992
KFC11	C. Tyler-Smith, Oxford	Cooper, 1992
alphoid	C. Tyler-Smith, Oxford	Tyler-Smith & Brown, 1987
TSPY	C. Tyler-Smith, Oxford	Tyler-Smith et al., 1988
52d	C. Tyler-Smith, Oxford	Vergnaud et al., 1986
50f2	J. Weissenbach, Paris	Vergnaud et al., 1986
AMY8e	C. Tyler-Smith, Oxford	Unpublished
692e		This work
752e		This work
757e	C. Tyler-Smith, Oxford	Unpublished data
758e	C. Tyler-Smith, Oxford	Unpublished data
AA15e	C. Tyler-Smith, Oxford	Cooper, 1992
AMELY	Oncor	Lau et al., 1989
GMGXY4	C. Sargent, Cambridge	Jones et al., 1994
GMGXY5	C. Sargent, Cambridge	Jones et al., 1994
GMGXY9	C. Sargent, Cambridge	Jones et al., 1994
GMGXY12	C. Sargent, Cambridge	Jones et al., 1994

2.3 Preparation of DNA

2.3.1 Preparation of plasmid DNA

2.3.1.a Solutions required

Solution I	Solution II	Solution III
50 mM glucose	0.2 M NaOH	5 M Potassium acetate
10 mM EDTA	1% SDS	
25 mM Tris-HCl, pH 8.0		

2.3.1.b Maxiprep of plasmid DNA

500 ml overnight culture of bacterial cells were grown. Plasmids were extracted based on the method of Birnboim and Doly (1979):

1. Pellet cells by centrifugation at 4500 rpm for 15 minutes
2. Pour off supernatant, and resuspend pellet in 25 ml solution I
3. Leave 5 minutes at room temperature
4. Add 50 ml solution II, and mix gently
5. Leave 5 minutes at room temperature
6. Add 37.5 ml solution III, and shake gently
7. Leave 5 minutes at room temperature
8. Centrifuge 15 minutes at 4000 rpm
9. Transfer supernatant to a fresh tube, and add 2 volumes ethanol.
10. Put on dry ice for 10 minutes
11. Centrifuge for 10 minutes, and discard supernatant. Resuspend pellet in 5 ml TE
12. Load onto caesium chloride / ethidium bromide gradient, as described in Maniatis et al. (1982) and extract plasmid DNA

Minipreps can be carried out using 1 ml of bacterial culture by scaling down the quantities of reagents involved, and omitting the caesium chloride step.

2.3.2 Preparation of PAC DNA

2 l of overnight culture of bacterial cells were grown, and the PAC DNA extracted with a QIAfilter plasmid Maxi kit (QIAGEN), according to the manufacturer's instructions.

2.3.3 Preparation of YAC DNA

2.3.3.a Solutions required

50 mM EDTA, pH 8.0

1 M sorbitol, 20 mM EDTA, 1 mg/ml zymolyase 20T (ICN biomedical)

1 M sorbitol, 20 mM EDTA, 14 mM β -mercaptoethanol, 1 mg/ml zymolyase

100 mM EDTA, 1% lithium dodecyl sulphate (LDS)

1% (w/v) LGT agarose (SeaPlaque, ICN), in 1 M sorbitol, 20 mM EDTA

2.3.2.b YAC miniprep

The method used is based on Southern et al. (1987), and may be scaled up as required.

1. Grow 10 ml overnight culture of yeast in SD medium
2. Centrifuge 10 minutes at 1500 rpm
3. Remove supernatant, and resuspend in 10 ml 50 mM EDTA
4. Spin for 10 minutes at 1500 rpm
5. Resuspend in 400 μ l sorbitol/EDTA/zymolyase
6. Add 500 μ l agarose at 37°C
7. Dispense into plug mould and set on ice for 10 minutes
8. Push plugs into 5 ml sorbitol/EDTA/ β -mercaptoethanol / zymolyase, and incubate at 37°C for 2 hours
9. Replace solution with 5 ml EDTA/LDS, and incubate at 50°C for 1 hour
10. Replace with fresh EDTA/LDS, and incubate overnight at 50°C
11. Plugs may be stored at room temperature for many months.

A list of the YAC clones mapped in this thesis is given in table 2.2 below.

Table 2.2: YAC clones mapped in this thesis				
YAC name	Probe content	Size (kb)	Chimaeric	Comments
152b4	14A3C	500	No	
416g2	109e	450	No	
SL147h5	219e	400	No	
711f4	RBF5	650	Yes	
900e7	RBF5	1115	Yes	
75g12	RBF5	380	No	
294g5	219e	385	No	Joins 14A3C and GMGY6 contigs
692b1	TSPY2, alphoid	890	Yes	
872e10	TSPY2, alphoid	1600	Yes	
752h1	TSPY3, AMY8e	1300	No	End-rescued; X-Y homologous probe
894f6	TSPY3, AMY8e	1110	Yes	
803d2	TSPY1	700	No	
759c10	752e, X-Y homologous probes	1135	No	
804f3	752e, X-Y homologous probes	1070	No	
847h2	752e, X-Y homologous probes	1160	Yes	
849g1	752e, X-Y homologous probes	950	No	
931d6	50f2/A, 758e	440	No	2 YACs in clone; Y chromosome YAC has lost right arm
948a2	50f2/A	800	Yes	
yOX81	50f2/A	260	No	
887g12	TSPY1, TSPY2	~2000	Yes	Fragmented and mapped. Joins TSPY and alphoid contigs. 1000 kb of Y sequence.

2.4 Manipulation of DNA

2.4.1 Restriction endonuclease digestion

2.4.1.a Solutions required

10 mg/ml BSA

0.1 mg/ml gelatin

Buffer M: 10 mM Tris-HCl, 10 mM MgCl₂, 50 mM NaCl, 1 mM DTT

Buffer H: 50 mM Tris-HCl, 10 mM MgCl₂, 0.1 M NaCl, 1 mM DTT

Buffer A: 20 mM Tris-acetate, 10 mM Mg acetate, 50 mM K acetate, 1 mM DTT

2.4.1.b Digestion in agarose plugs

Gel plugs containing DNA were equilibrated 2 x 30 minutes in TE, followed by 2 x 30 minutes in the appropriate buffer. Then the enzyme was added at the desired concentration in a total volume of 100 µl containing 1 x buffer, 1 x gelatin, and, if necessary, 1 x BSA. Partial digests were carried out for 2 hours at the specified temperature for each enzyme (although a maximum temperature of 50°C was used in order to avoid melting the plugs); complete digests were typically performed overnight. Digestion was stopped by equilibration for 1 hour in TE, on ice.

2.4.1 c Digestion of DNA in solution

An appropriate volume of enzyme, buffer, gelatin, and spermidine were added to the DNA solution, and incubated according to the enzyme supplier's recommendations.

2.5 Gel electrophoresis

2.5.1 Solutions and reagents

50 x TAE buffer: 2 M Tris-acetate pH 7.5, 50 mM EDTA

Gels: Conventional gels: 1.0% (w/v) agarose (Sigma) in 0.5 x TAE

Pulsed field: 1.5% (w/v) agarose (Sigma) in 0.5 x TAE

Preparative gels: 1.5% (w/v) SeaPlaque (ICN Biomedicals) in 0.5 x TAE

PCR products: 3% (w/v) Nu Sieve; 1% (w/v) agarose (Sigma) in 0.5 x TAE

Loading buffer: 30% Ficoll, Orange G

5 µg/ml ethidium bromide

2.5.2 Conventional gels

These were run using a minigel box from Pharmacia, with 1% agarose / 0.5 x TAE at 100 V for c. 1 hour. Staining was carried out with ethidium bromide for 15 minutes, followed by destaining in deionized water for 20 minutes. Conventional gels were run with *Hind*III-digested λ DNA as a marker.

2.5.3 Preparative gels

These were run in a Pharmacia minigel box with 1.5% SeaPlaque gel, under conditions as above. After staining with ethidium bromide, the desired band was excised and placed in a sterile plastic tube for further manipulation.

2.5.4 Purification of DNA from agarose gels

0.5 ml of DNA in agarose was placed in a clean eppendorf tube, incubated at 65°C for 10 minutes, then at 37°C for 5 minutes, mixing gently to ensure all agarose has melted. Then 1 unit of agarase (New England Biolabs) was added, and incubated at 37°C for 2 hours. DNA was then phenol / chloroform extracted, followed by ethanol precipitation, and resuspension in TE buffer. DNA was then stored at -20°C. Alternatively, a GeneClean kit (Bio 101) was used according to the supplier's instructions.

2.5.5 Pulsed-field gel electrophoresis

Pulsed field gels were run on the 'Waltzer' apparatus, as described by Southern et al. (1987). 200 ml of 1.5% agarose in TAE was poured into a gel mould, and allowed to set. The wells were loaded with DNA in gel plugs, and sealed with a small amount of 0.5% agarose. Pulsed-field gels were run with concatemers of λ DNA as a marker (obtained from New England Biolabs), giving markers of multiples of 48.5 kb. Gels were run at 150 V, with buffer being circulated at 17.5°C. The run time varied from 26 to 30 hours, and the pulse time varied from 25 seconds to 70 seconds, in order to separate fragments up to a maximum of 500 kb or 1200 kb respectively. Staining was carried out with ethidium bromide for 30 minutes, followed by 60 minutes destaining in water.

2.5.6 Gel viewing and photography

Ethidium bromide stained gels were viewed on a UV transilluminator ($\lambda_{\text{max}} = 300$ nm), and photographed through a red filter with Polaroid 667 film.

2.6 Southern analysis

2.6.1 Blotting of gels

The blotting apparatus was set up as described in Southern (1975), modified as follows. Hybond N+ nylon filters were used. Gels containing small DNA molecules (i.e., conventional gels) were placed directly on the apparatus; larger DNA molecules (i.e., pulsed field gels) were first depurinated for 15 minutes in 0.25 M HCl, and then neutralised for 30 minutes in 1.5 M NaCl, 0.4 M NaOH, before being placed on the blotting apparatus. Blotting was performed with 0.4 M NaOH overnight, and the

filters were rinsed with 6 x SSC after being removed from the apparatus. If storage was necessary, the filters were sealed with cling film and stored at -20°C.

2.6.2 Hybridisation of filters

2.6.2.a Solutions and reagents

Hybridisation mix: 10% dextran sulphate, 0.375 M Na₂HPO₄, 0.125 M NaH₂PO₄, 5 x Denhart's solution (1 x Denhart's solution is 0.02% bovine serum albumin, 0.02% ficoll, 0.02% polyvinylpyrrolidone), 1% SDS

Low-temperature hybridisation mix: 50% formamide, 5 x Denhart's solution, 6 x SSC, 0.5% SDS, 5% dextran sulphate

Salmon sperm DNA (0.1 mg/ml)

α-32P dCTP (Amersham)

Wash mix: 1% SDS, 0.1 x SSC (high stringency)

1% SDS, 2 x SSC (low stringency)

Solution O: 1.25 M Tris-HCl, 0.125 M MgCl₂, pH 8.0

Solution A: 1 ml solution O + 18 µl β-mercaptoethanol + 5 µl 0.1 M dATP + 5 µl 0.1 M dTTP + 5 µl 0.1 M dGTP

Solution B: 2 M Hepes titrated to pH 6.6 with 4 M NaOH

Solution C: Hexadeoxyribonucleotides (Boehringer) evenly suspended in TE at 90 OD units / ml

Solution OLB: Mix A: B: C in the ratio of 100:250:150

2.6.2.b Hybridisation and washing of filters

The filter was placed in a tube containing 20 ml hybridisation mix, filtered through a 0.45 µm filter (Millipore), plus 200 µl boiled salmon sperm DNA. This was placed in a 68°C incubator (Techne) for 1-2 hours and rotated horizontally. The radiolabelled probe (see below) was then boiled for 3 minutes, quenched on ice, and added to the mix. The hybridisation was kept at 68°C overnight. Library screening was performed at 42°C overnight, using 50 ml of low-temperature hybridisation mix for 24 filters (the complete CEPH library).

Unhybridised probe was removed by washing the filter 2 x 10 minutes followed by 2 x 30 minutes with wash mix (the stringency of the wash required depending on the

application and the properties of the probe itself) at 65°C. The filter was then removed from the hybridisation tube and sealed in cling film, before being visualised. If it was necessary to rehybridise a filter, it was first stripped by immersing in boiling 0.5% SDS, and allowing to cool.

2.6.2.c Radiolabelling of DNA probes

Around 50-100 ng of DNA was diluted with TE to a volume of 35 µl, and boiled for 3 minutes. The following components were then added to an eppendorf tube on ice:

10 µl OLB

2 µl 10 mg/ml BSA

35 µl DNA

2 units Klenow fragment *E. coli* DNA polymerase I

4 µl α -³²P dCTP

The reaction was kept at room temperature for 2 hours. Incorporation of labelled nucleotides was, if desired, checked by TCA precipitation.

2.6.2.d Removal of unincorporated nucleotides

Unincorporated nucleotides were separated by spun-column chromatography.

1. Plug the bottom of a 1 ml syringe with glass wool
2. Fill the syringe with Sephadex G25 equilibrated in TE buffer.
3. Spin in a centrifuge for 5 minutes at 1500 rpm
4. Repeat steps 2 and 3 until around 0.9 ml of the syringe is filled
5. Spin twice with TE
6. Columns may be stored at 4°C for several months.

To separate the unincorporated nucleotides:

1. Place an eppendorf tube at the mouth of the syringe
2. Add the reaction mix to the top
3. Spin for 5 minutes at 1500 rpm
4. The eppendorf contains the labelled probe

2.6.2.e Imaging and Autoradiography

Sealed filters were first imaged in a computerised InstantImager, from Packard. After gaining a first picture of the filter, it was then placed in a cassette with Fuji X-ray film at -70°C for a time depending on the intensity of the signal, up to 2 weeks. The film was developed according to the manufacturer's instructions.

2.7 Plasmid end-rescue of probes from YACs

2.7.1 Preparation of electrocompetent cells

XL1-Blue *E. coli* cells were prepared as described (Dower, Miller & Ragsdale, 1988), and stored in liquid nitrogen until needed for use.

2.7.2 Preparation of DNA

Six agarose plugs containing YAC DNA were digested to completion with an appropriate restriction enzyme (*Nde*I, *Xho*I, or *Xho*I / *Sal*I), and the fragments separated on a 1% agarose gel. The desired size fraction was excised from the gel, and the DNA isolated with the GeneClean kit.

The DNA solution was then self-ligated overnight at 4°C with 10 units of T4 ligase (Boehringer), in a total volume of 500 µl. After ethanol precipitation, the ligated DNA was resuspended in 10 µl water.

2.7.3 Electroporation

2 µl of DNA was added to 40 µl of thawed competent cells. Using a BioRad GenePulser, a pulse of 2.48 kV was applied to the cells. Immediately, 1 ml of YT-broth was added, and the cells incubated at 37°C for 1 hour. After this time, the cells were plated onto selective YT-agar, and incubated overnight at 37°C. Separate colonies were then picked, and the transformed *E. coli* colony purified. The insert DNA was isolated as described in section 2.3.1

2.8 Transformation of yeast cells

2.8.1 Reagents and solutions

TE / Lithium Acetate, prepared fresh from 10 x TE and 1M LiAc, pH 7.5 stocks

YPD medium (1% yeast extract, 2% peptone, 2% glucose)

40% PEG 4000, 1 x TE, 0.1 M LiAc

0.1 mg/ml salmon sperm DNA

2.8.2 Transformation of yeast

The procedure followed was that of Gietz et al., 1992.

1. Grow yeast to 1×10^7 cells / ml in YPD medium.
2. Harvest and wash cells in sterile H₂O, resuspend in 1 ml sterile H₂O, transfer to 1.5 ml Eppendorf tubes, and pellet cells.
3. Wash cells in 1 ml TE / LiAc, and resuspend cells to 2×10^9 cells / ml in TE / LiAc.
4. Mix 50 μ l of yeast cell suspension with 1 μ g transforming DNA and 50 μ g salmon sperm DNA.
5. Add 300 μ l sterile PEG / TE / LiAc solution, and mix well.
6. Incubate at 30°C with agitation for 30'.
7. Place in a 42°C waterbath for exactly 15'.
8. Spin down in microfuge, and resuspend pellet in 1 ml of TE
9. Plate out onto appropriate selective medium.

2.9 Polymerase chain reaction (PCR)

2.9.1 Reagents and solutions

PCR buffer: 50 mM KCl, 10 mM Tris-HCl pH 8.4, 0.01% gelatin

MgCl₂, 25 mM stock

dNTP mix (100 x): mix equal quantities of each of the four 100 mM dNTPs (Pharmacia). This gives a 100 x stock (25 mM)

Taq polymerase: 2.5 units/ μ l (Promega)

2.9.2 Primers

Table 2.3: PCR primers and conditions used in this work					
Name	Sequence (5' - 3')	Product size	Conditions	Reference	Comments
YACL1	GTG TGG TCG CCA TGA TCG CG	330 bp	92°C 30'' 55°C 30'' 72°C 60'' 30 cycles 2.5 mM MgCl ₂	Hirst et al., 1991	To amplify hybridisation probes from the YAC vector left and right arms
YACL2	ATG CGG TAG TTT ATC ACA GTT AA				
YACR1	GAT CAT CGT CGC GCT CCA GCG AAA GC	265 bp			
YACR2	CTC GCC ACT TCG GGC TCA				
Alu 559	AAG TCG CGG CTT GCA GTG AGC CGA GAT	Undefined – a mixture of products; YAC- specific	94°C 1' 55°C 45'' 68°C 5' 35 cycles 1.5 mM MgCl ₂	Nelson et al., 1991	For amplifying Alu-vector products, to use in hybridisation to detect overlaps
Alu 278	CCG AAT TCG CCT CCC AAA GTG CTG GGA TTA CAG				
YAC4L	CGG AAT TCG CCA AGT TGG TTT AAG GCG CAA GAC				
YAC4R	GGA AGC TTG GCG AGT CGA ACG CCC GAT CTC AAG				
TTY1A	CTC TGG GAA TCA AAT TCG AGG	82 bp	92°C 30'' 58°C 30'' 72°C 1' 35 cycles 1.5 mM MgCl ₂	Lahn & Page, 1997	See discussion of these genes in section 1.2.1 and 1.2.2
TTY1B	GTC TTT CAG CCA ATC CAA GG				
TTY2A	GAC AAC TCT GAC AGC CAG G	88 bp			
TTY2B	GTC AGA ACT CCC AAA CAG G				
UTYA	GCA TCA TAA TAT GGA TCT AGT AGG	65 bp			
UTYB	GGA GAT ACT GAA TAG CAT AGC				
DBYA	TTG CAG CAA AGT CAC CCT TA	257 bp			
DBYB	CAC TAA AGC AGG AGG GGA AG				
DFFRYA	GAC ACA TGC AGG GTG AAG TG	226 bp			
DFFRYB	AAC TGC TCA AAG CAG GCA AT				
PRYA	CTT GAG TGA GGC AGT GAC CA	121 bp			
PRYB	CAC TCA ACG CCT CTC CTT TC				
TB4YA	TCT TGC CCG ATT TTA ACC AC	115 bp			
TB4YB	AGG AAT TTG ACC CCT GTG AA				
BPY1A	GCA GCA ACT AAG TTT AGG GC	78 bp			
BPY1B	TTG TCA TCA ACA TGG GAA GC				

FBY1for	GAC CTT TTC TTC ACG TGA C	578 bp	94°C 1' 58°C 1' 72°C 45" 30 cycles 1.5 mM MgCl ₂	Rappold et al., 1995	Amplifies sequence from the <i>PRKY</i> gene
FBY1rev	AAA ACA GAC AAC ATA AAA TTA CA				
92r7A	TGC ATG AAC ACA AAA GAC GTA	55 bp	94°C 20" 60°C 20" 72°C 30" 35 cycles 1.5 mM MgCl ₂	Zerjal, pers. comm.	55 bp product can be digested with <i>Hind</i> III when C allele present; not when T allele present.
92r7B	GCA TTG TTA AAT ATG ACC AGC				
SRY-1532A	TCC TTA GCA ACC ATT AAT CTG G	167 bp	94°C 20" 60°C 20" 72°C 30" 35 cycles 1.5 mM MgCl ₂	Kwok et al., 1996	167 bp product digests with <i>Dra</i> III when G allele present, but not when A allele present.
SRY-1532B	AAA TAG CAA AAA ATG ACA CAA GGC				
50f2 A	CTC AAG CTA GGA CAA AGG GAA AAG G	196 bp 50f2C locus; 1.7 – 2.7 kb 50f2B locus	95°C 1' 66°C 4' 26 cycles	Jobling & Tyler-Smith, 1995	Amplifies two 50f2 loci, B and C.
50f2 B	GAG GTA GAT GCT GAA GCG GTA TAG				
YAP1	CAG GGG AAG ATA AAG AAA TA	455 bp if YAP+; 150 bp if YAP-	94°C 20" 60°C 20" 72°C 30" 35 cycles 1.5 mM MgCl ₂	Hammer & Horai, 1995	Amplifies one band of differing size depending on the presence or absence of an Alu insertion.
YAP2	ACT GCT AAA AGG GGA TGG				

2.9.3 PCR

PCR was performed in a MJ Peltier thermal cycler, according to conditions described in the above table. Products were then size fractionated on agarose gels; the concentration of agarose depending on the expected product sizes.

2.10 Fluorescent in-situ hybridisation (FISH)

2.10.1 Preparation of metaphase spreads

2.10.1.a Solutions and reagents

PBS (phosphate buffered saline): 0.9 mM CaCl_2 , 2.7 mM KCl, 1.5 mM KH_2PO_4 , 0.5 mM MgSO_4 , 137 mM NaCl, 4.3 mM Na_2HPO_4 ; pH 7.6

Colcemid: 10 $\mu\text{g/ml}$ stock solution (Gibco BRL)

Hypotonic solution: 40 mM KCl, 0.5 mM EDTA, 20 mM Hepes pH 7.4

Fixative: methanol: acetic acid 3:1

45% (v/v) acetic acid

2.10.1.b Preparation of slides

1. Grow the cells to 70% saturation in 15 ml medium. Add 300 μl Colcemid solution, and leave at 37°C for 75 minutes.
2. Wash cells twice in PBS, and add 5 ml of prewarmed (37°C) hypotonic solution. Leave at 37°C for 25 minutes, then pellet in a centrifuge for 5 minutes at 500 rpm.
3. Resuspend in a few drops of ice cold fixative, and slowly add 1 ml fixative. Keep on ice for 20 minutes, then slowly top up with 12 ml fixative, and leave at -20°C overnight.
4. Pellet at 300 rpm for 5 minutes, and wash in 10 ml fixative. Repeat three times, and finally resuspend in 1 ml fixative.
5. Drop 45% acetic acid onto a clean glass microscope slide, then one drop of cell suspension. Leave to air dry, then view under a microscope to check spreading of cells.
6. Slides may be stored in a sealed box at -20°C

2.10.2 Preparation of extended chromatin slides

The protocol followed is that of Haaf and Ward (1994)

2.10.2.a Solutions and reagents

PBS (as above)

Hypotonic solution: 10 mM Hepes (pH 7.3), 30 mM glycerol, 1.0 mM CaCl_2 , 0.8 mM MgCl_2

Fixative: Methanol, chilled to -20°C

2.10.2.b Preparation of slides

Grow cells in RPMI medium, as above. Wash twice in PBS, then resuspend to a density of $10^3 - 10^5$ cells / ml in hypotonic solution. Leave in hypotonic for 10' at room temperature, then centrifuge 0.5 ml of cell suspension onto a glass microscope slide for 4' at 800 rpm. The slides are then fixed in cold methanol for 30'.

2.10.3 Probe labelling

Probes were labelled by nick translation with the BioNick™ labelling system (Gibco BRL). This kit is designed for labelling with biotin; it was also modified for use with digoxigenin.

2.10.3.a Solutions and reagents

10 x enzyme mix:	0.5 units/ μ l DNA polymerase I
	0.0075 units/ μ l DNase I
	50 mM Tris-HCl
	5 mM Mg acetate
	1 mM β mercaptoethanol
	0.1 mM phenylmethanesulphonyl fluoride
	50% (v/v) glycerol
	100 μ g/ml nuclease-free BSA
10 x dNTP mix:	0.2 mM each dCTP, dGTP, dTTP
	0.1 mM dATP
	0.1 mM biotin-14-dUTP or digoxigenin-11-dUTP
	500 mM Tris-HCl (pH 7.8)
	50 mM MgCl ₂
	100 mM β mercaptoethanol
	100 μ g/ml nuclease-free BSA
Stop buffer:	300 mM EDTA

2.10.3.b Labelling of probes

1. Add the following components to an eppendorf tube on ice:
 - 5 μ l 10 x dNTP mix
 - 1 μ g DNA
 - H₂O to 40 μ l
 - 5 μ l 10 x enzyme mix
2. Mix well, centrifuge briefly, then incubate at 16°C for 1 hour
3. Add 5 μ l stop buffer
4. Purify labelled probe by ethanol precipitation.
5. Store probes in TE at -20°C.

2.10.4 FISH

2.10.4.a Solutions and reagents

Cot-1 DNA 1 mg/ml (Gibco BRL)

Wash buffer: 4 x SSC, 0.05% Tween-20

Hybrisol VIII (Oncor)

2 x SSC

70% formamide (v/v) in 2 x SSC

50% formamide (v/v) in 2 x SSC

Ice cold 70%, 80%, 100% (v/v) ethanol

Labelled probes

FITC – avidin (Oncor)

Rhodamine – anti-digoxigenin (Oncor)

DAPI in antifade (Oncor)

2.10.4.b Hybridisation

1. Wash slides in 2 x SSC at 37°C for 1 hour
2. Dehydrate slides for 2 minutes each in cold 70%, 80%, 100% (v/v) ethanol
3. Denature DNA for 2 minutes in 70% (v/v) formamide in 2 x SSC
4. Dehydrate in an ethanol series as above
5. Probe preparation:
 - i. Add 20-100 ng labelled probe to 30 μ l Hybrisol
 - ii. Denature for 5 minutes at 70°C

iii. Leave to reanneal for 30 minutes at 37°C

6. Add probe to slide. Seal with a glass coverslip and rubber glue.

7. Leave overnight in a moist chamber at 37°C

2.10.4.c Washes and detection

1. Remove rubber glue, and soak slide in 4 x SSC to remove coverslip

2. Wash slide in 50% formamide, 2 x SSC at 42°C for 10 minutes, twice.

3. Wash in 2 x SSC twice, for 10 minutes, at 37°C

4. Rinse once in cold 2 x SSC, then once in PBD, for 2 minutes.

For biotin-labelled probes:

5. Add 60 µl FITC – avidin, and seal with a plastic coverslip.

6. Incubate in a moist chamber at 37°C for 20 minutes.

7. Wash 4 times in PBD, for 2 minutes each

For digoxigenin-labelled probes:

8. Add 60 µl rhodamine – anti-digoxigenin, and seal with a plastic coverslip.

9. Incubate in a moist chamber at 37°C for 20 minutes.

10. Wash 4 times in PBD, for 2 minutes each

To visualise DNA:

11. Add 20 µl DAPI in antifade, and seal with a glass coverslip and rubber glue.

2.10.4.d Microscopy and image processing

Slides were examined with a Zeiss Axioskop fluorescence microscope, with separate filters for visualising DAPI, FITC, and rhodamine. Images were acquired with a CCD camera, and coloured and merged using an Apple Macintosh running IPLab Spectrum image processing software.

Chapter 3: Mapping of the Y chromosome long arm

3.1 Contigs 14A3C and GMGY6

These two contigs are considered together. They lie in the distal region of the area being considered in this work, around 2 Mb from the alphoid DNA of the centromere. 14A3C is proximal to GMGY6. At the outset of this work, the 14A3C contig covered around 900 kb, extending from the 14A3C probe to the probe 219e, an end-rescued fragment of YAC yOX 219. This is made clearer by referring to the contig map given in figure 3.11 below. The whole of this contig consisted of five YACs from a single library, OXEN (a library made from a 49, XYYYYY individual). The GMGY6 contig covered 1300 kb, consisting of 6 OXEN YACs and a single CEPH library YAC (787g2). No probes were available at the extremities of the contig; RBF5 lay some 300 kb from the proximal end, while 92e was found 540 kb from the distal end of the contig. It was known that the two contigs were closely linked, Southern analysis of genomic DNA suggesting a gap between the probes 219e and RBF5 of not more than 500 kb (C. Tyler-Smith, pers. comm.)

The first mapping experiments took place with a set of 3 CEPH YACs (152b4, 416g2, 711f4) and one from the St. Louis YAC library (SL147h5), known from previous work by other groups (Affara et al., 1994) to contain several probes from these two contigs. The first two YACs were a kind gift from Peter Vogt, Heidelberg; while the latter two were supplied by Nabeel Affara, Cambridge. The actual probe content is listed in table 3.1, below (data not shown).

Table 3.1: Probe content of initial 14A3C and GMGY6 YACs							
	14A3C	109e	116e	219e	RBF5	GMGY6	92e
152b4	+	+	+	-	-	-	-
416g2	-	+	+	-	-	-	-
SL147h5	-	-	+	+	-	-	-
711f4	-	-	-	-	+	+	+

For the locations of the probes, please see the map at the end of this section (Figure 3.11).

As can be immediately seen, the YACs fall into two sets: 711f4 lies on the GMGY6 contig, while the other three all overlap to some extent, and lie on the 14A3C contig. Although the two sets do not obviously overlap, it was hoped that they might extend the coverage of the region to some extent.

The mapping process used for all YACs in this work will be described in some detail here, in order that it may be made familiar; it will not be mentioned in other sections. The initial step after identifying YACs of interest is to ensure the preparations used in the actual mapping are pure clones, and so contain only one YAC per cell. Variant forms may exist for a number of reasons: firstly, the problems with instability discussed in section 1.8.2 can lead to alternative forms of the YAC being found in any culture of cells. Secondly, during the transformation process used to create the library, the host cell may have taken up more than one YAC. This can lead to problems in trying to map only one of the YACs in the cell; and in addition, recombination can occur between the different YACs, leading to further rearrangement of the cloned DNA.

In order to avoid these problems, a number of clones are taken of the YAC of interest, and run on a pulsed-field gel, along with the original stock, in order to identify the YAC. If there is more than one YAC band in a strain, the gel itself can be hybridised with the probe of interest in order to determine which YAC corresponds with the desired probe. The yeast clone can then be grown without selection for a number of generations, followed by the reapplication of selection pressure for a YAC; the lack of selection initially allows cells to lose one or both of the YACs, while when selection is reapplied, only those which have retained at least one of the YACs will survive. The yeast strain can then be reanalysed to determine if the appropriate YAC has been lost from the strain.

If each clone contains only one YAC, but these are of different sizes, then it is most likely that the YAC itself is unstable, and prone to deletion over the generations. In this case, it is assumed that the largest form of the YAC is the complete form, and this clone is chosen to be used in the mapping. This can also be checked by Southern blotting followed by hybridisation.

Once the desired clone is prepared, the mapping is carried out by means of partial digests. In this work, five restriction enzymes are used to make the map: *Bsi*WI, *Bss*HII, *Mlu*I, *Sac*II, and *Sfi*I. Several samples of each YAC are digested to varying extents of completion (most conveniently by varying the concentration of enzyme used in each reaction), and the DNA is then separated on a pulsed-field gel. If the restriction digestion is judged to have worked, then the gel is then transferred to a nylon membrane by Southern blotting, which is then probed sequentially with the left and the right vector arms of the YAC. Autoradiograms made from the hybridised membrane will then show the locations of the restriction sites along the YAC.

A consensus map is assembled from the partial digest data, and this map then allows the YAC to be located on the main map of the region. With small YACs, complete mapping data is obtained from each end, while larger YACs do not usually provide a complete map with either end. If the measured positions of the sites differ, a judgement is made about which is the more accurate, usually the smaller. Other discrepancies which may arise in mapping are resolved similarly. It is assumed that measured fragment sizes may vary between maps, while the order and number of restriction sites is less likely to be in error. However, in certain cases this data may conflict – for instance, two sites close together may only be resolved as a single band on one gel, and as two bands on another; or two sites thought to coincide on one gel may be seen as being slightly offset on another. It is assumed that data from nearer the ends of the YAC are more reliable than data from the centre of the YAC; and data from smaller fragments will be more accurate than that from larger fragments. In addition to this, once the map has been made, a judgement must be made regarding allowable discrepancies between maps, and chimaeric portions of the YAC. It is usually clear if part of a map is chimaeric, principally from the order and number of restriction sites, rather than the fragment sizes. The mapping then proceeds to another round of YAC selection and mapping.

The mapping data from the above-mentioned four YACs is presented below.

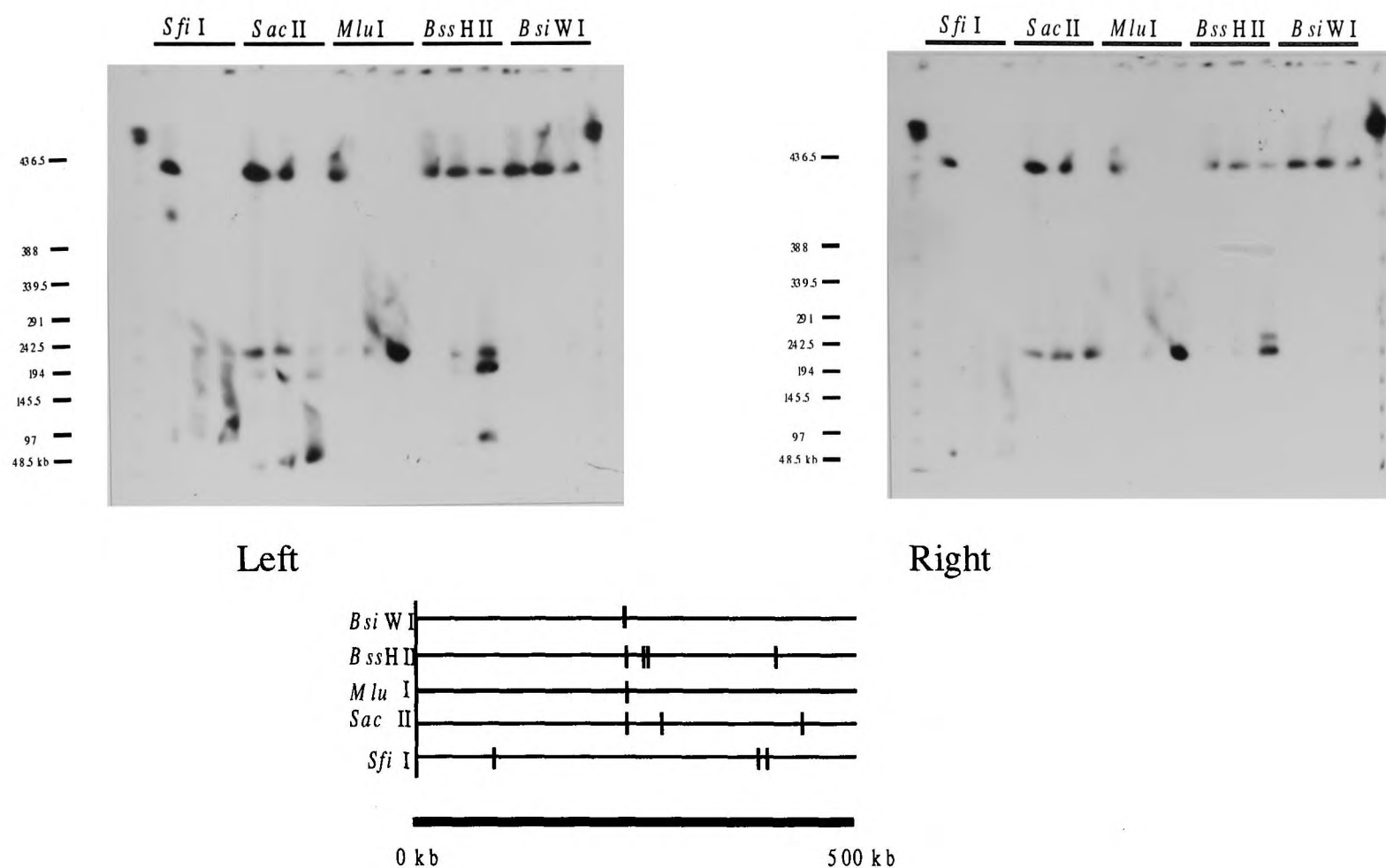


Figure 3.1: Mapping data of YAC 152b4. The left arm of the YAC is on the left of the figure, in both the partial digests and the map.

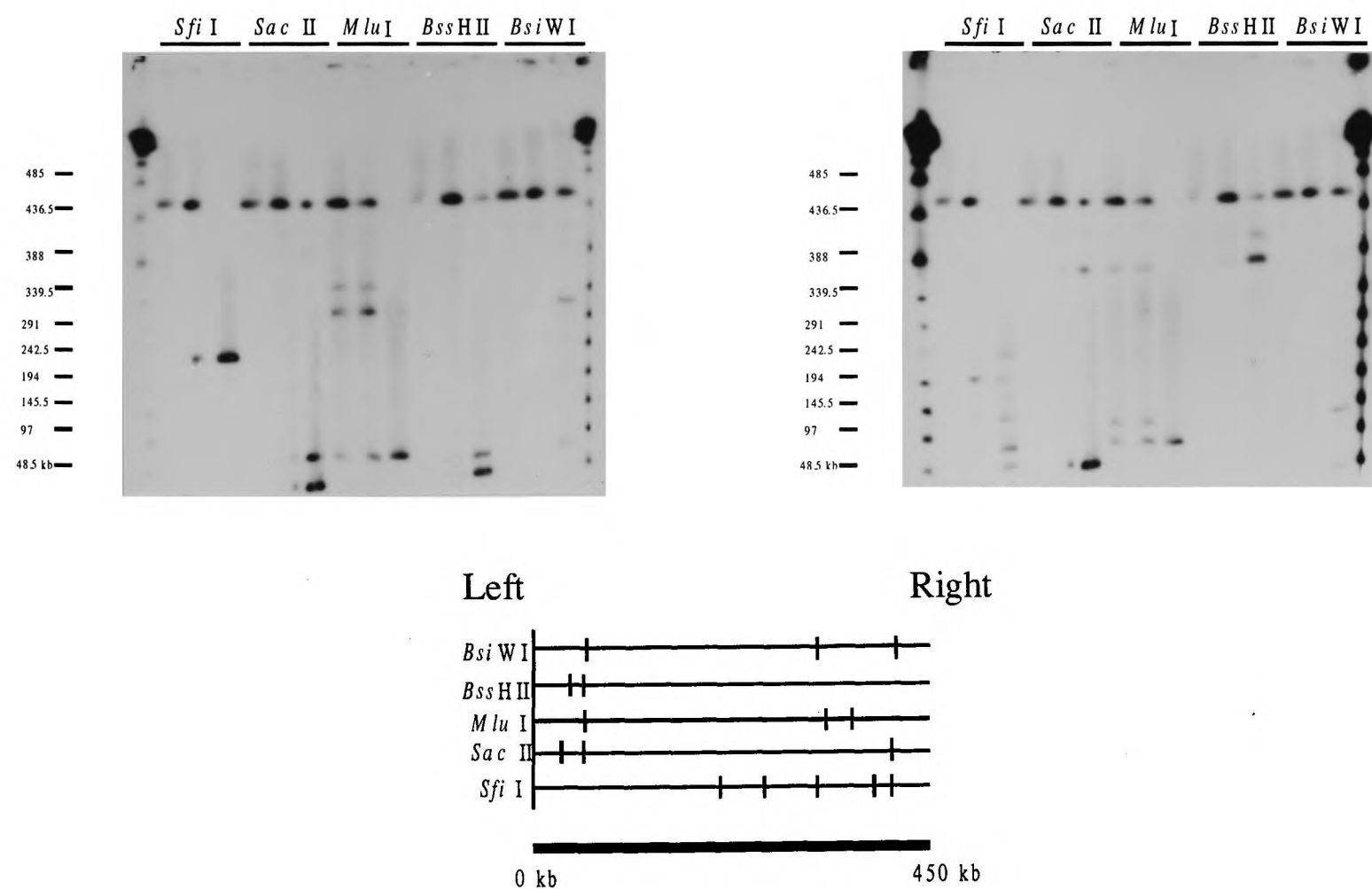


Figure 3.2: Mapping data of YAC 416g2. The left arm of the YAC is on the left-hand side of the page, in both the partial digests and the consensus map.

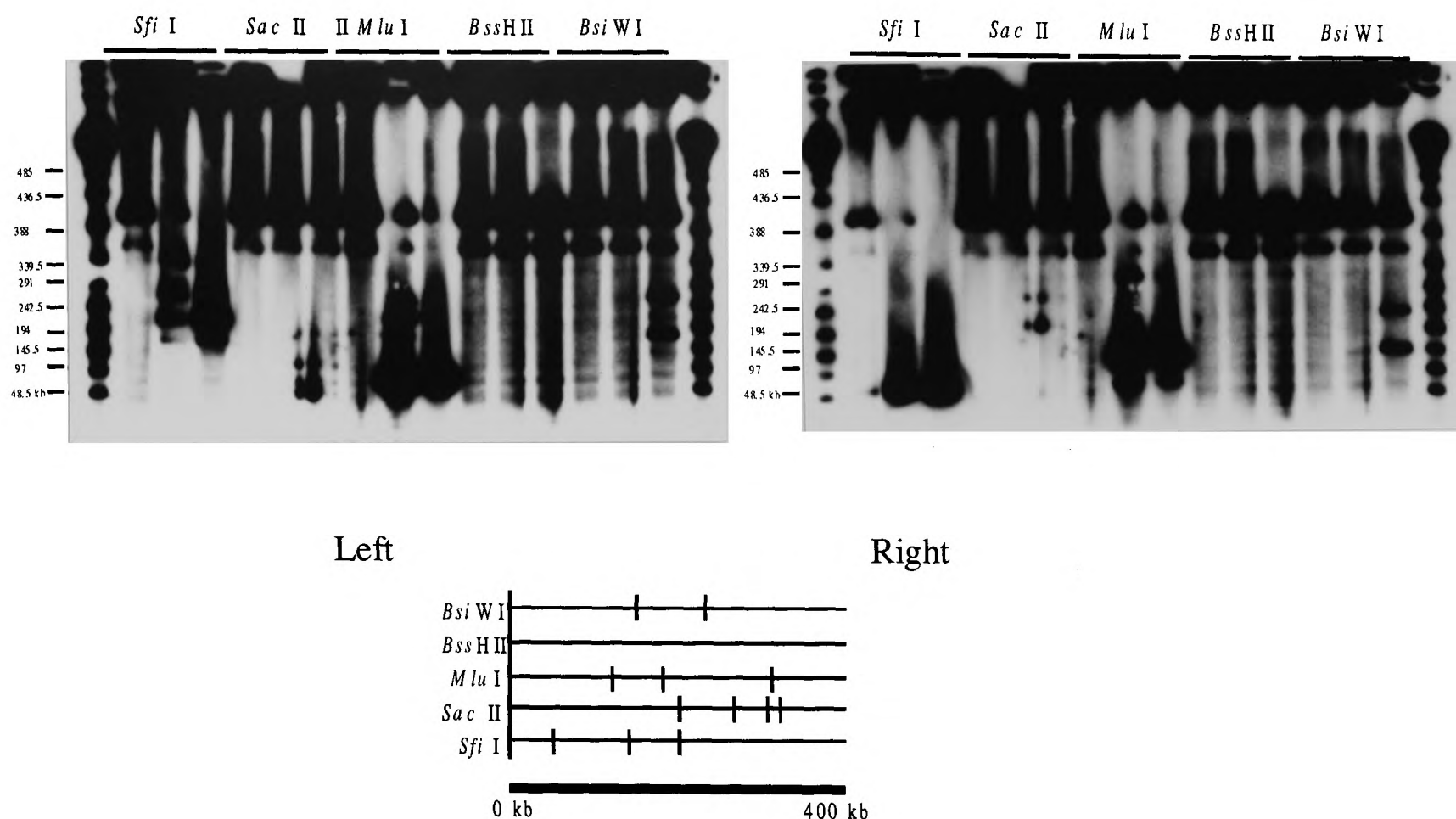


Figure 3.3: Mapping data of St. Louis library YAC SL147h5. The left-hand side of the page corresponds with the left arm of the YAC.

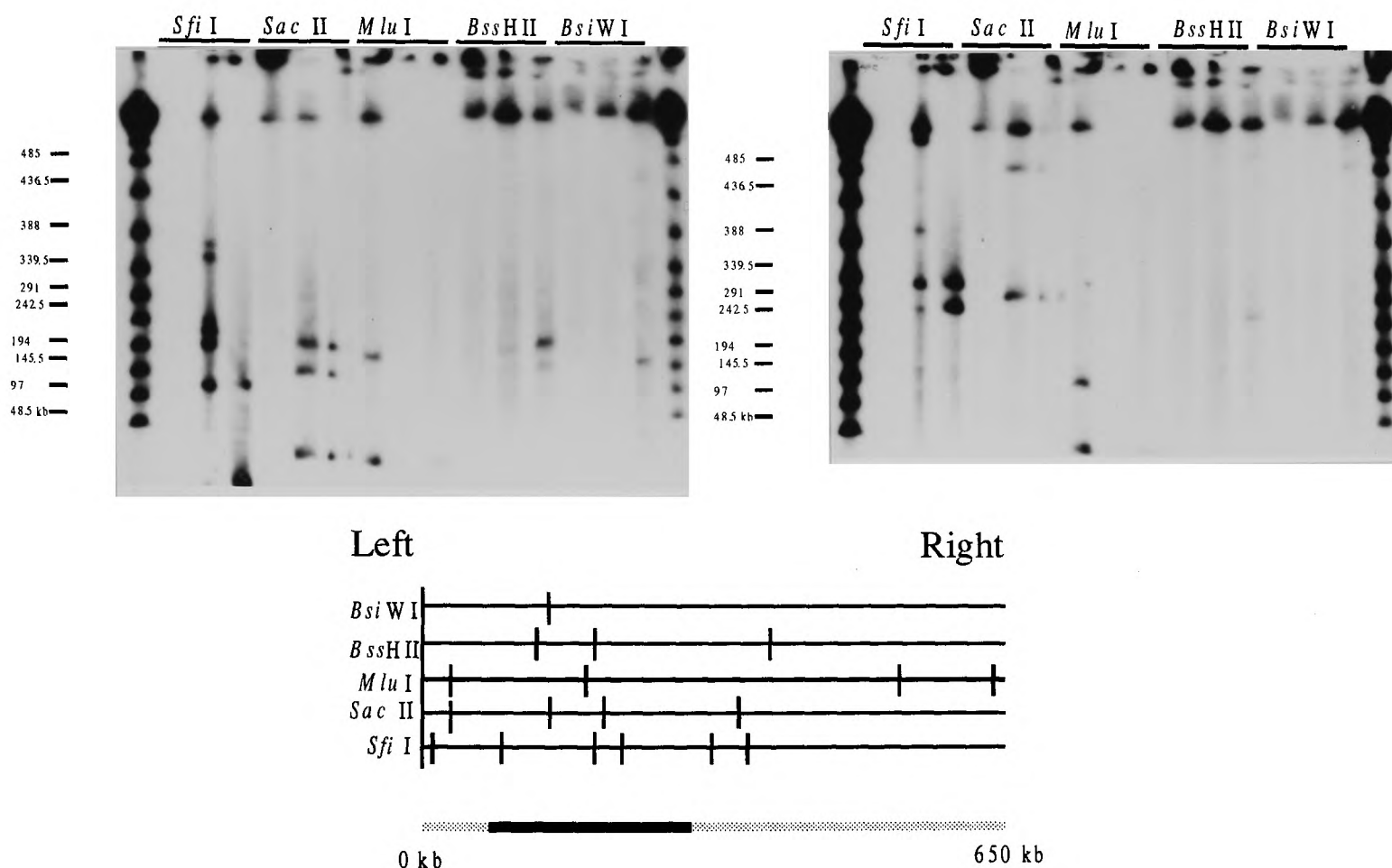


Figure 3.4: Map of YAC 711f4. The left-hand side of the page corresponds with the left arm of the YAC. The shaded portions of the bar beneath the map represent the chimaeric regions of the YAC; the black portion represents the Y-chromosomal sequences.

The YACs shown in figures 3.1 – 3.3 (152b4, 416g2, and SL147h5) could all be placed on the map of contig 14A3C without any difficulties. One anomaly was noted in the map of 416g2, which appeared to contain two *Sfi*I sites not previously noted on the contig map. One of these sites was in the region of overlap with 152b4; and this YAC did not appear to contain this additional restriction site. Accordingly, this was interpreted as a minor difference in this one YAC, rather than a signal that these YACs were from different genomic regions. Another problem was noted with YAC SL147h5 showing an additional large band in almost all tracks (around 350 kb) – as this was in most lanes, and the same size with both the left and right vector arm probes, it was concluded that this was a deleted form of the YAC present in the DNA preparation, and discounted from the mapping process.

The fourth YAC mapped here, 711f4, shown in figure 3.4, was slightly more complicated. Despite being known to contain the probes RBF5 and GMGY6, the map did not immediately appear to correspond with the contig map. However, it was noted that one small region of around 200 kb in the interior of the YAC showed identity with the small region of the contig map containing both these probes. Hence it was decided that this YAC was chimaeric, with only a small region being derived from this region of the Y chromosome.

The placing of these YACs on the contig maps can be seen in figure 3.11. Note that in all the contig maps in this and the following chapter, the left arm of the YACs is denoted by a box marked on the YAC; solid boxes designate those YACs from which end-rescued probes have been produced. Chimaeric parts of YACs are marked by zigzag lines, rather than shading as is the case in the individual YAC figures.

After having mapped these four YACs, the map still had not been extended by any distance. Accordingly, it was decided to screen the entire CEPH mega-YAC library.

The CEPH mega-YAC library was screened with a cocktail of probes from throughout the region of interest on both the long and short arms of the chromosome. The data from the library screening is not shown here, as only a single YAC is described; a fuller description of the library screening and interpretation of results is given in chapter 4. A number of YACs were picked from the library, and digests of

these were run on conventional electrophoresis gels, to be reprobbed with the desired sequences, so confirming any positive results. Figure 3.5 shows a *Hind*III digest being probed with the RBF5 probe.

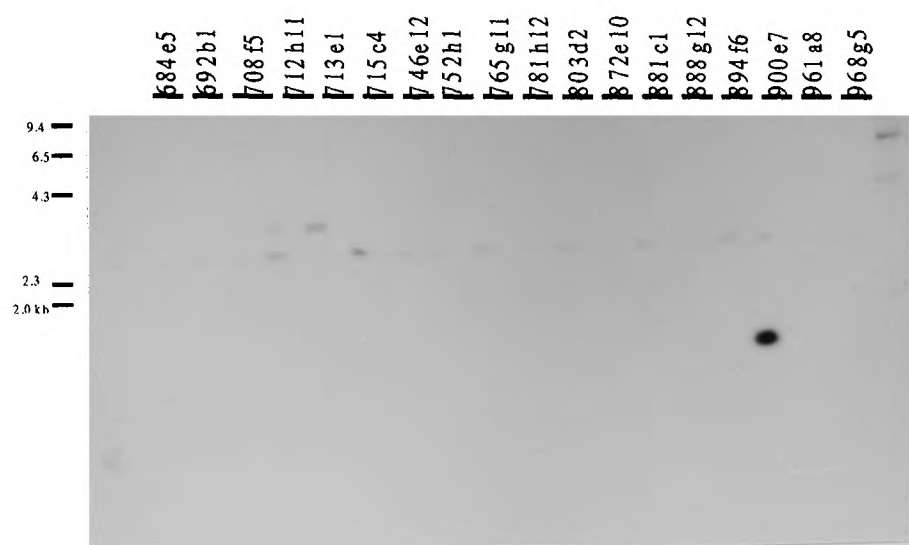


Figure 3.5: Assorted YACs were digested with *Hind*III, fractionated by gel electrophoresis, transferred to a nylon membrane, and probed with the RBF5 sequence.

As is apparent, only a single YAC from this round of screening contained the RBF5 sequence. Accordingly, YAC 900e7 was colony-purified and mapped. The data is shown in figure 3.6 below.

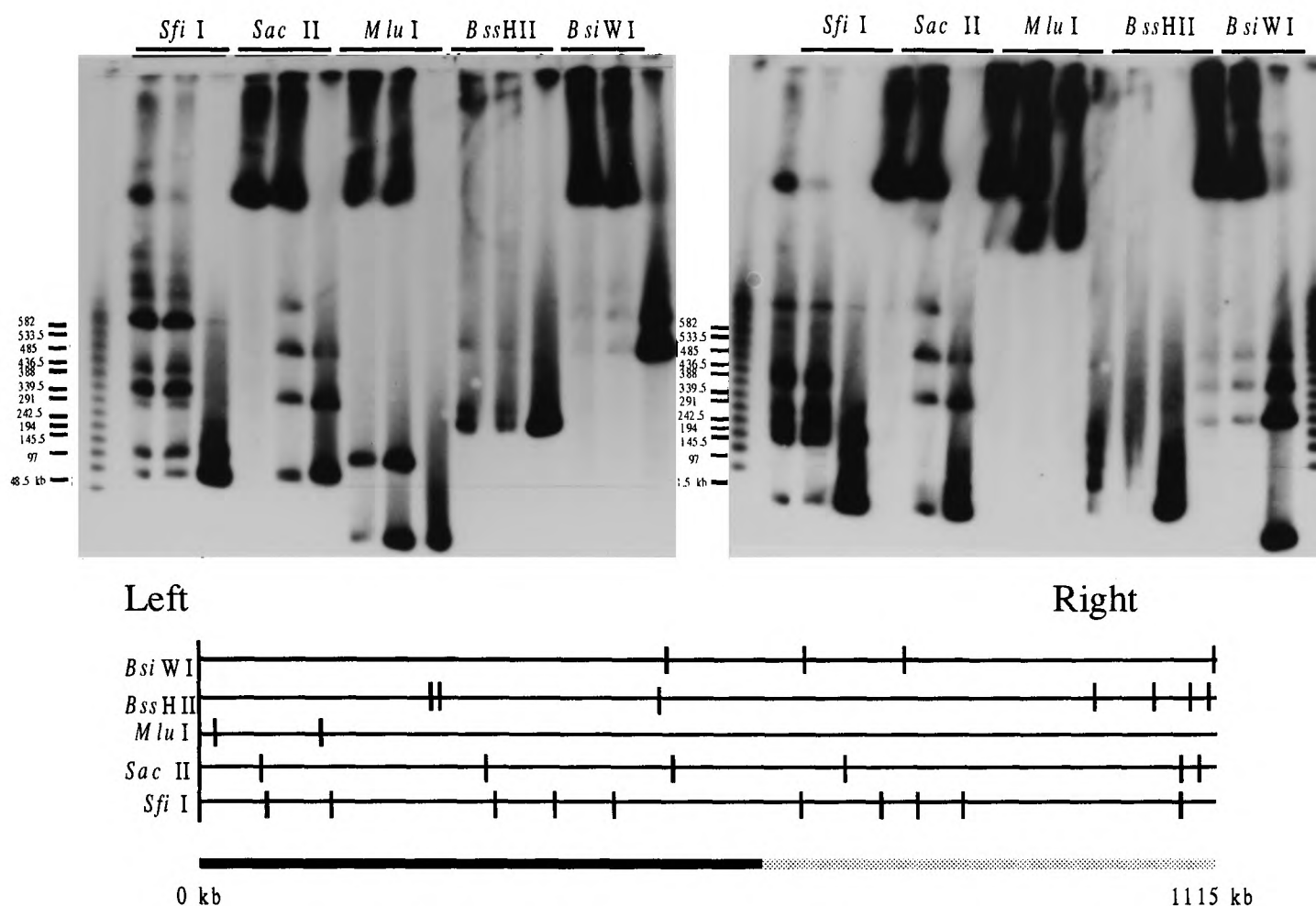


Figure 3.6: Mapping of YAC 900e7. The left-hand side of the page relates to the left arm of the YAC, in both the photographs and the map. The shaded part of the bar beneath the map marks the chimaeric region of the YAC; the black part of the bar marks the Y-chromosomal region of the YAC.

This YAC, 900e7, was also found to be chimaeric, and also went no further toward closing the gap between the 14A3C and GMGY6 contigs. Of course, it must be emphasised that these mapping efforts are not wasted, as every YAC mapped helps to confirm the consensus contig map; these YACs particularly so, as the bulk of these contigs consisted solely of OXEN YACs, which may have shown some differences from the CEPH YAC library.

During the course of this work, it was observed from the results shown in Affara et al. (1996) that a number of other CEPH YACs had been localised in this region. Accordingly, the YACs 75g12, 294g5, 334c7 and 410d2 were obtained from D. Le Paslier at CEPH in Paris. The results of probing these YACs with 219e and RBF5 (from either side of the gap between the two contigs) are shown in figure 3.7.

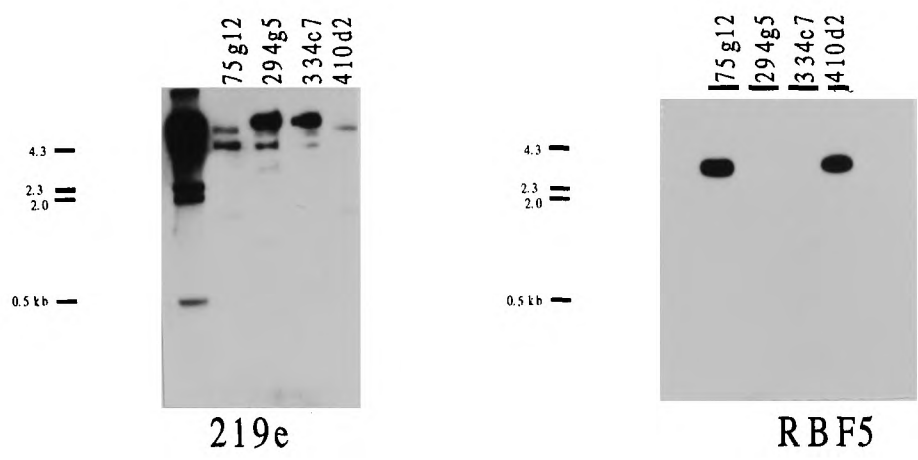


Figure 3.7: *Hind*III digests of the YACs obtained from CEPH hybridised with a) 219e, and b) RBF5

As can be seen from the figure, none of the YACs contained both 219e and RBF5, so it was not apparent whether any of the YACs would close the gap between the contigs. It was decided to map both 75g12 and 294g5 initially, to determine if the contigs were extended by these YACs. The mapping data is given in figures 3.8 (75g12) and 3.9 (294g5).

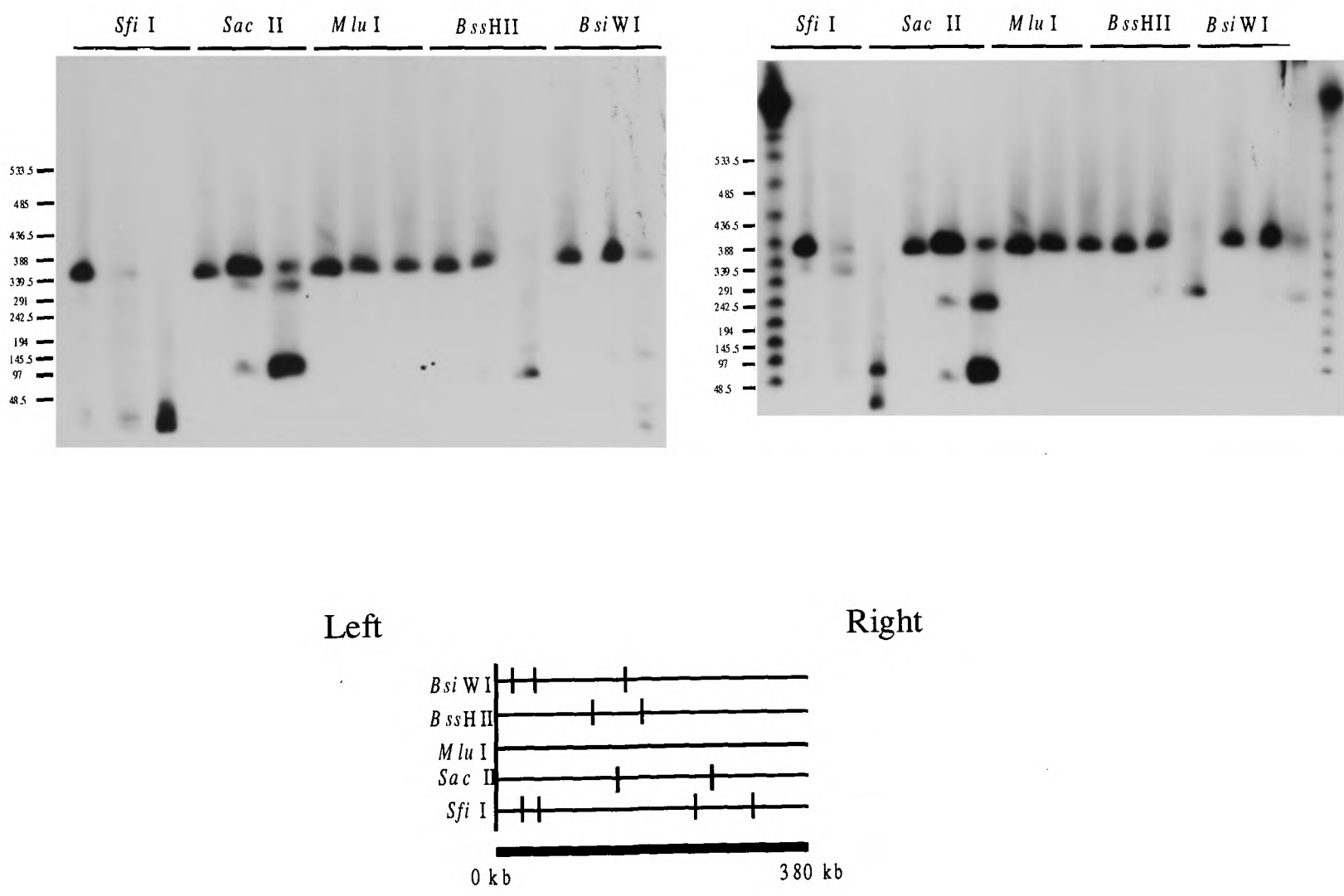


Figure 3.8: Mapping data from YAC 75g12. No *Mlu*I data were obtained.

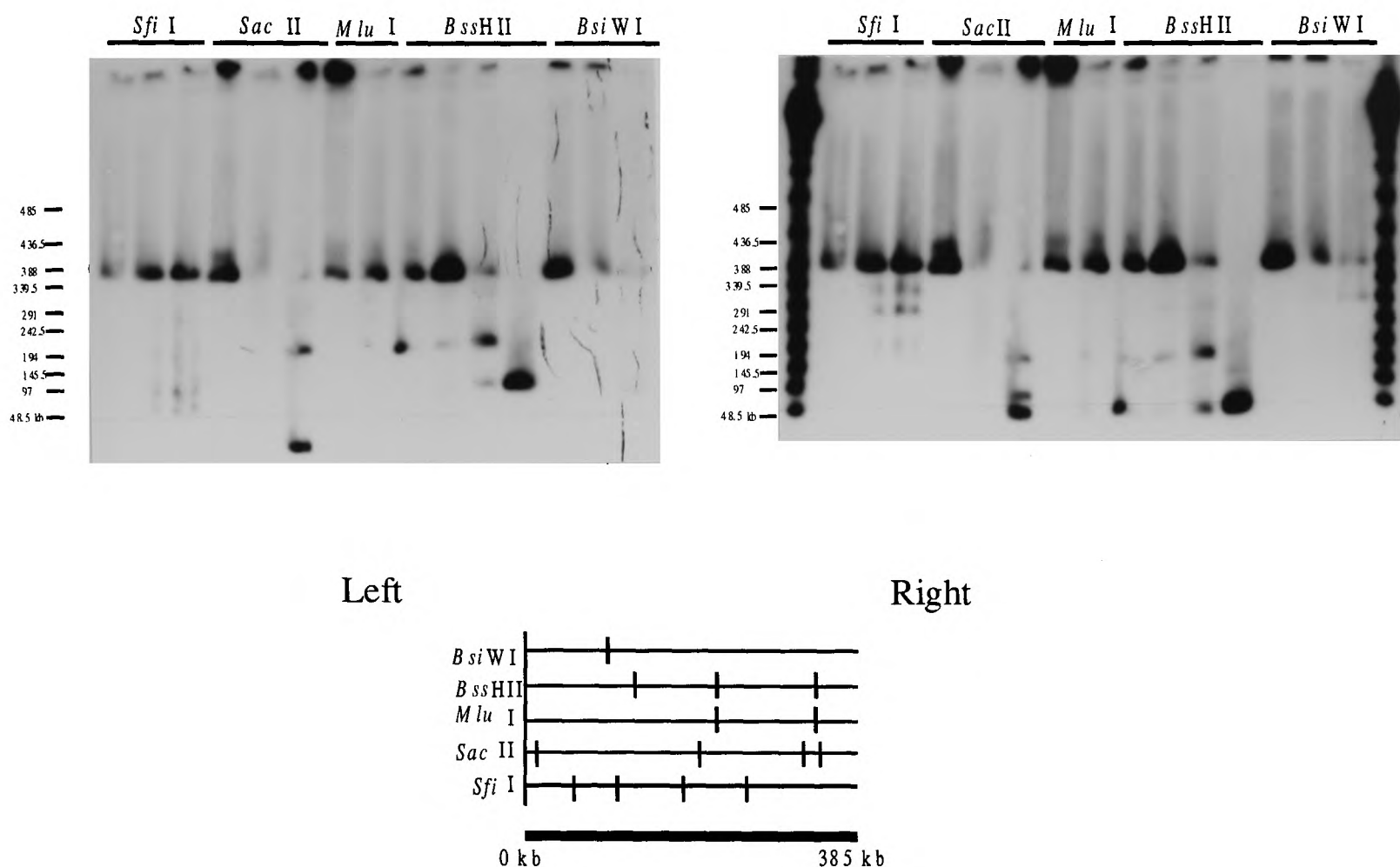


Figure 3.9: Mapping data of YAC 294g5.

Studying the map of 75g12 showed that it lay entirely on the GMGY6 contig, and did not extend the coverage by any distance. YAC 294g5, however, was found to extend the 14A3C contig by some 200 kb beyond the previous most distal point. The map at this point appeared to correspond with the map of the GMGY6 contig on the other side of the gap. Alu-PCR (see Nelson et al., 1991; and section 2.9) was used to amplify a fragment from the left end of this YAC, and hybridise it to a panel of YACs both from the area of supposed overlap and from elsewhere (see figure 3.10 below – yOX 219 was not expected to overlap the 294-left probe, while the other YACs tested were). The data confirmed that the gap between the two contigs had been closed, so creating a single contig covering some 2300 kb of the Y chromosomal long arm. The final contig map is shown in figure 3.11.

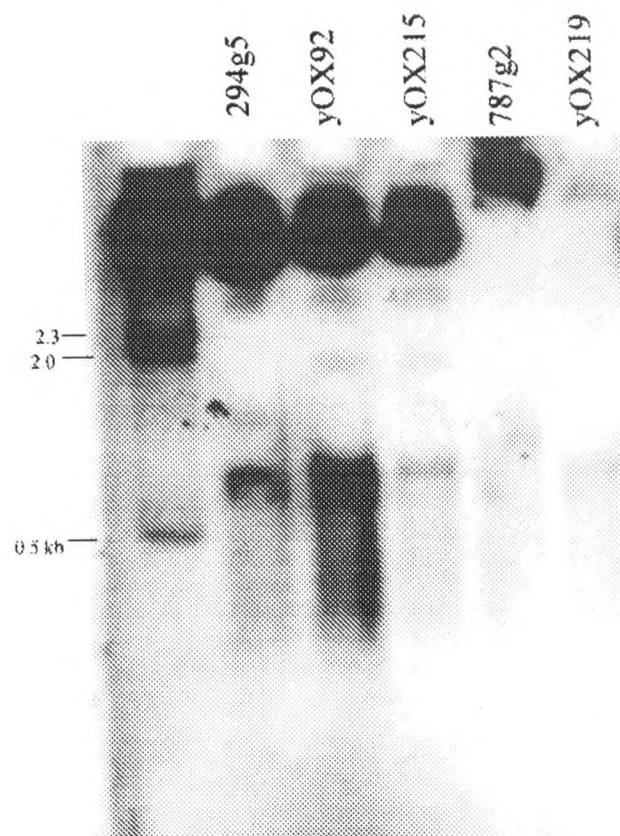


Figure 3.10: Hybridisation of a panel of *HindIII*-digested YACs to the Alu-vector PCR product 294-left. The YAC 787g2 did not digest completely.

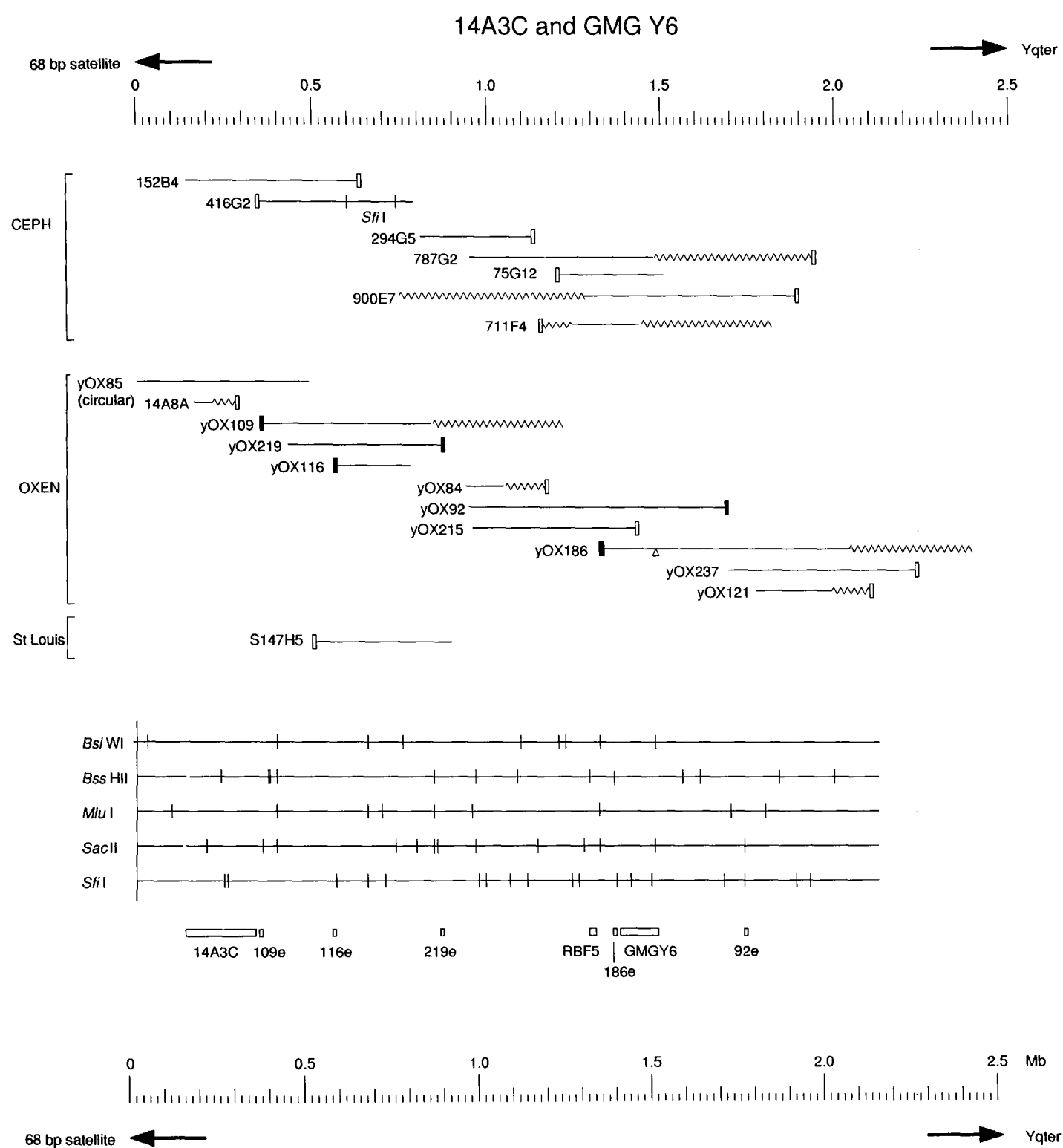


Figure 3.11: The final consensus map of the 14A3C and GMGY6 contig. YACs are shown as straight lines; zigzags denote chimaeric regions of the YAC; boxes mark the left end of the YAC; filled boxes signify the existence of an end-rescued probe derived from the YAC.

3.2 The 68 bp satellite contig

This contig was the most troublesome to study of all six contigs in this work. The original contig (as mapped by Cooper, 1992) is composed of three OXEN YACs, covering a total of only 230 kb. The content of the YACs is almost entirely satellite or repeated sequences – the 5 bp, 48 bp, and 68 bp satellites are all present in this contig, along with the probes KFC52, KFC43, and KFC11, which all detect numerous loci on genomic DNA (see Cooper, Fisher & Tyler-Smith, 1992). Additionally, no unique sequences are known to be present on this contig.

The lack of unique sequences, and the abundance of repetitive, non-Y specific sequences in the contig mean that it is extremely difficult to study in cloned DNA. Not only are repeated sequences unstable in YACs, so may be under-represented in the libraries, but also if the repeats were used to screen the library, then many clones from throughout the genome would be identified, with no immediate means of distinguishing Y chromosomal from non-Y chromosomal clones.

One potential avenue of study is opened up by the knowledge of the long-range genomic map of this region (Cooper, 1992). It is known that this contig is linked on genomic DNA to the alphoid and 14A3C contigs, on the 48 bp and 68 bp sides, respectively. The size of these gaps has been measured as being 0.7 Mb to the alphoid contig, and 1 Mb to the 14A3C contig. The vast majority of these intervening sequences are thought to consist of further satellite and non-satellite repeats. However, one probe, 64a, is known to lie within 200 kb of the 48 bp satellite repeat. This 64a probe is not unique, but present in two copies in the genome, one autosomal and one Y chromosomal, which are distinguishable from one another. Thus, it is theoretically possible to use this as a probe to isolate clones from a library for mapping.

Both the CEPH and OXEN libraries were screened with the probe 64a, and no clones were obtained, whether autosomal or Y chromosomal. A previous screening of the smaller CEPH library (clone numbers below 600; as opposed to the mega-YAC library used in this thesis) found a single 64a autosomal clone, but no Y chromosomal

clones (Cooper, 1992). This supports previous theories that the 64a region is grossly under-represented in YAC clones, possibly due to the difficulty of stable maintenance of a low copy-number repeat in the midst of highly repeated sequences in the yeast host. Thus, it does not seem likely that this contig would be able to be extended by means of YACs.

Two PACs were then obtained, a kind gift from Stefan Kirsch, Heidelberg, which did contain the 64a probe. The PACs, 64aI and 64aII were digested with *EcoRI*, separated on a conventional gel, transferred to a nylon membrane by Southern blotting, and probed with various of the probes from the 68 bp satellite contig, in order to determine if the clones did link 64a with the YAC contig. The results of the probing, with 64a, 68 bp satellite, KFC52, and KFC11, are shown in figure 3.12 below.

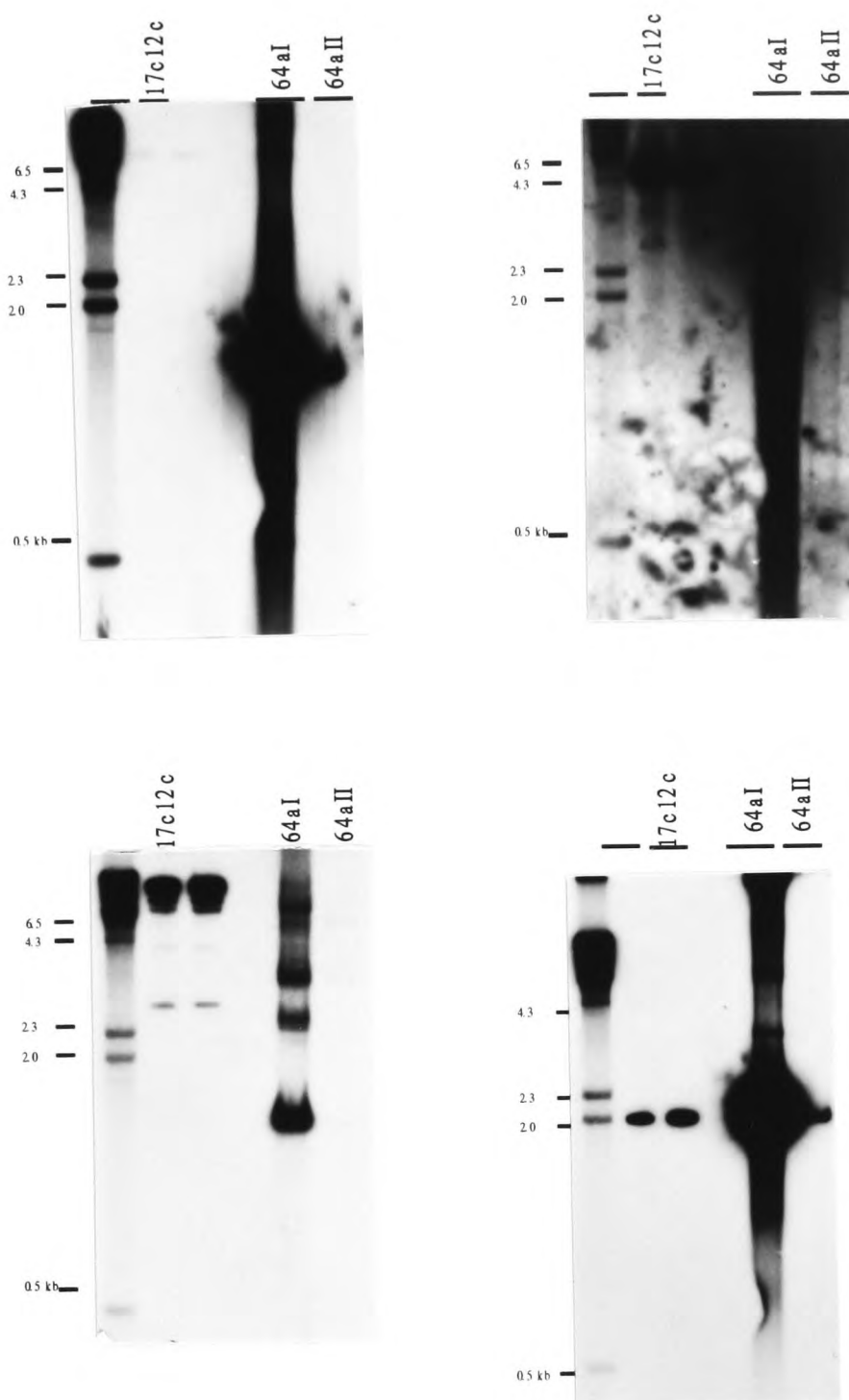


Figure 3.12: Probing of *Eco*RI digests of the two 64a PACs with, from left to right, (top row) 64a and KFC52, and (bottom row) 68 bp satellite, and KFC11. See text for details.

These figures show a number of things. First of all, the PACs do contain the 64a probe, as well as the two KFC probes (KFC52 and KFC11). Both these probes are from the same region, around the block of 48 bp satellite of the YAC contig. The lane marked 17C12C on the figure contains the YAC on which the contig is based, so confirming this interpretation of the results. The final probe used above was the 68 bp satellite, which, although present in both the YAC and the PAC 64aI, is present in different forms. This suggests that the PAC does not extend this far along the contig, although it does contain some amount of satellite sequence elsewhere along its length.

As there exist two separate PACs, both of which appear to link 64a with the YAC contig, it seems unlikely that the link is just the result of some chance rearrangement. Also, as both PACs extend as far as the 64a region, but 64aI also contains some 68 bp satellite sequence, we can tentatively place this second satellite block proximal to 64a, toward the centromere. However, the decision was taken not to proceed further with mapping of the PACs or searching for more PACs to cover the gaps. This region of the chromosome had already been mapped on genomic DNA (Cooper, Fisher & Tyler-Smith, 1992), and was known to consist largely of satellite sequences; so it was not judged pressing to complete the map in cloned DNA.

This data does, however, confirm that it is possible to cover regions which are highly-repeated or unclonable in YACs in an alternative cloning system, in this case PACs. This approach to closing the gaps in the map is discussed further in section 7.1. For completeness, the current map of the YAC contig is given below, in figure 3.13.

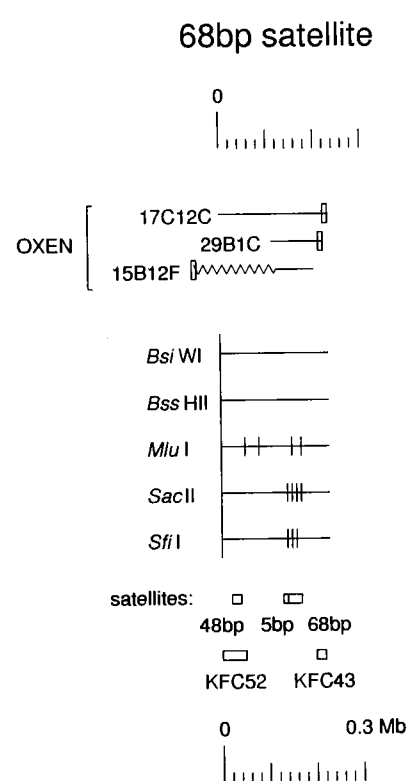
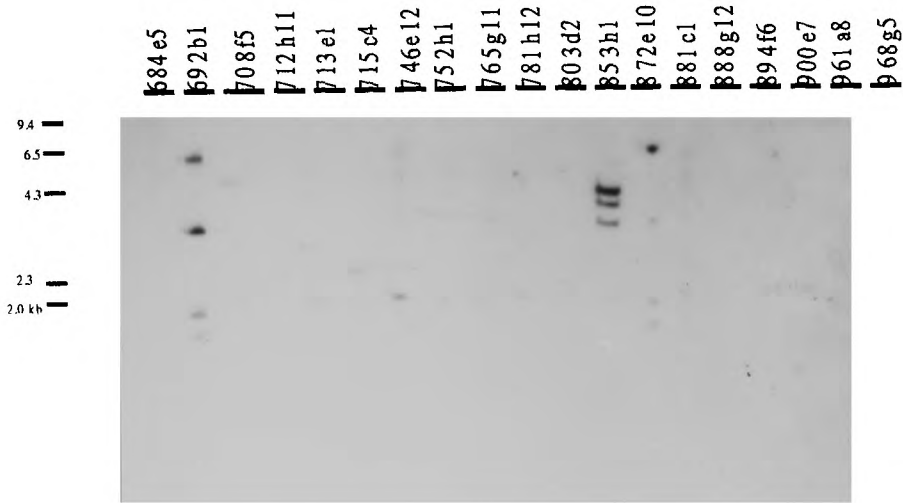


Figure 3.13: Consensus map of the 68bp satellite contig. (From Cooper, 1992)

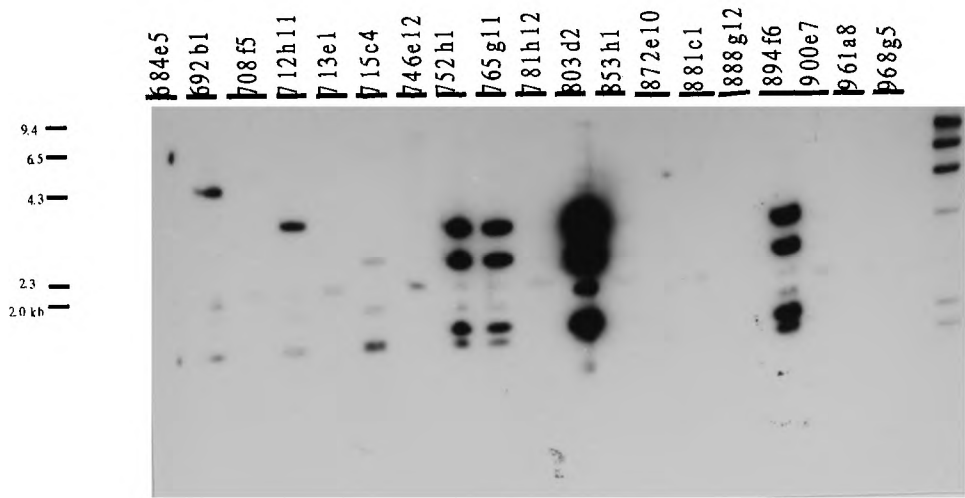
4.1: The alphoid contig

At the outset of this work, the alphoid contig consisted of 9 OXEN YACs, 3 St. Louis YACs, and 3 CEPH YACs. The contig covered 700 kb of DNA, on the short arm side of the alphoid array. Many other YACs had been found entirely within the alphoid array, although these could not be aligned across this array, due to its highly repeated nature. Furthermore, the size of this array is polymorphic between individuals. So, the contig is not contiguous, but is as complete as the technique allows; indeed it would perhaps be impossible to assemble a complete contig covering the whole of the alphoid array. Also the general problems inherent in mapping large repeated blocks prevent an accurate contig being assembled. However, YACs from both the OXEN and CEPH libraries were found which cover the edges of the alphoid array, and the transition to other sequences. YACs linked the centromeric alphoid array with flanking sequences of 5 bp satellite on the Yq side, and a number of other repeated sequences exist on the Yp side of the alphoid array. These repeated sequences included the TSPY2 block (also known as TSPY or Y-190 minor); and copies of the 52d and 50f2 repeats, which are found in several locations over the Y chromosome. For the map of this contig, please see figure 4.5 in this section.

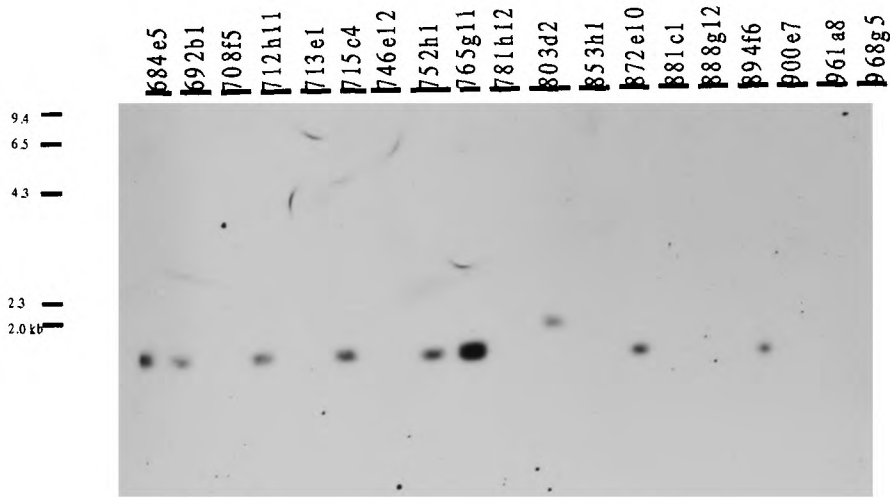
The present work commenced by screening the entire CEPH mega-YAC library with a cocktail of probes, including 52d and 50f2 from this contig. The alphoid probe was not used in this initial screening, due to its abundance in the genome, but was used in secondary screening of the selected clones. 18 YACs were picked out (this includes YACs which had been selected by hybridisation with other probes in the cocktail, not originating on this contig). These YACs were digested with *HindIII* (for alphoid and TSPY) or *EcoRI* (for 52d, 50f2, and AMY8e), transferred to a nylon membrane by Southern blotting, and hybridised with the probes used in the initial library screening. The results of the hybridisation with probes alphoid, 52d, TSPY, 50f2 and AMY8e (see section 4.3, The *AMELY* contig for the location of this probe) are shown in figure 4.1 below. The results are then summarised in table 4.1.



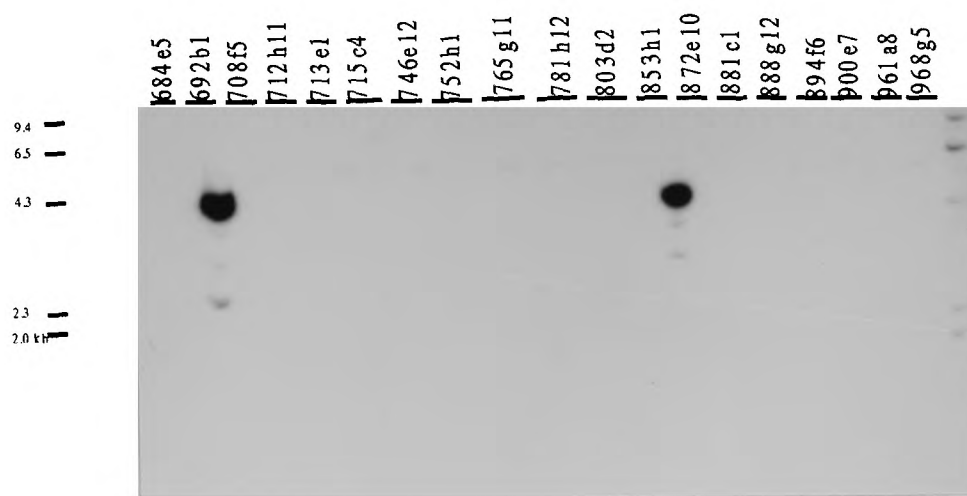
a) Alphoid



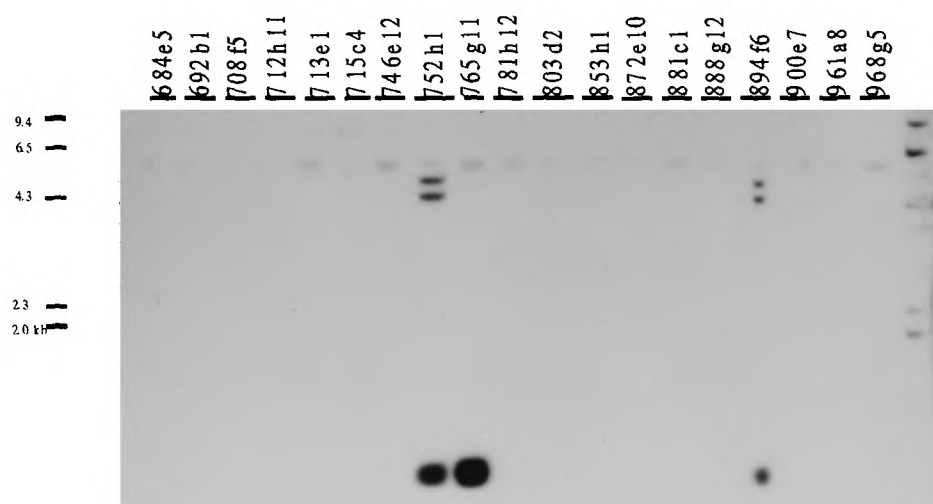
b) TSPY



c) 52d



d) 50f2



e) AMY8e

Figure 4.1: CEPH YACs digested with *Hind*III [a) and b)] or *Eco*RI [c), d), and e)], and hybridised with, from a) to e), alphoid, TSPY, 52d, 50f2, and AMY8e.

Table 4.1: Results of probing CEPH YACs in figure 4.1					
	AMY8e	TSPY	52d	50f2	Alphoid
684e5	-	-	C	-	-
692b1	-	+	C	D	+
708f5	-	-	-	-	-
712h11	-	+	C	-	-
713e1	-	-	-	-	-
715c4	-	-	C	-	-
746e12	-	-	-	-	-
752h1	+	+	C	-	-
765g11	+	+	C	-	-
781h12	-	-	-	-	-
803d2	-	+	B	-	-
853h1	-	-	-	-	+
872e10	-	+	C	D	+
881c1	-	-	-	-	-
888g12	-	-	-	-	-
894f6	-	+	C	-	-
900e7	-	-	-	-	-
961a8	-	-	-	-	-
968g5	-	-	-	-	-

In the above table, a ‘+’ or a ‘-’ for the probes AMY8e, TSPY, and alphoid mean that the YAC named was, respectively, positive or negative when probed. Probes 52d and 50f2 are present in the genome in a number of copies, given the letters A – E. The table gives the copy of each probe which was present in the YAC being tested.

Three YACs from this group contained some alphoid sequences, and two of them – 692b1 and 872e10 – contained other Y chromosomal sequences. Accordingly, these were mapped (see figures 4.2 and 4.3). It will be noted that these two YACs also contained the adjacent sequences TSPY, 50f2/D and 52d/C. Likewise, these three sequences were shared by the YACs 752h1, 765g11 and 894f6, which also contained the probe AMY8e from the *AMELY* contig. This raised the possibility that these YACs might link the alphoid contig to the *AMELY* contig – more details on the mapping of these YACs will be found in section 4.3. The final YAC it is worth pointing out here is 803d2, which contained a copy of the 52d probe; this time the B copy, rather than the C copy found on the alphoid contig. 52d/B can be found on the TSPY contig, discussed in section 4.2.

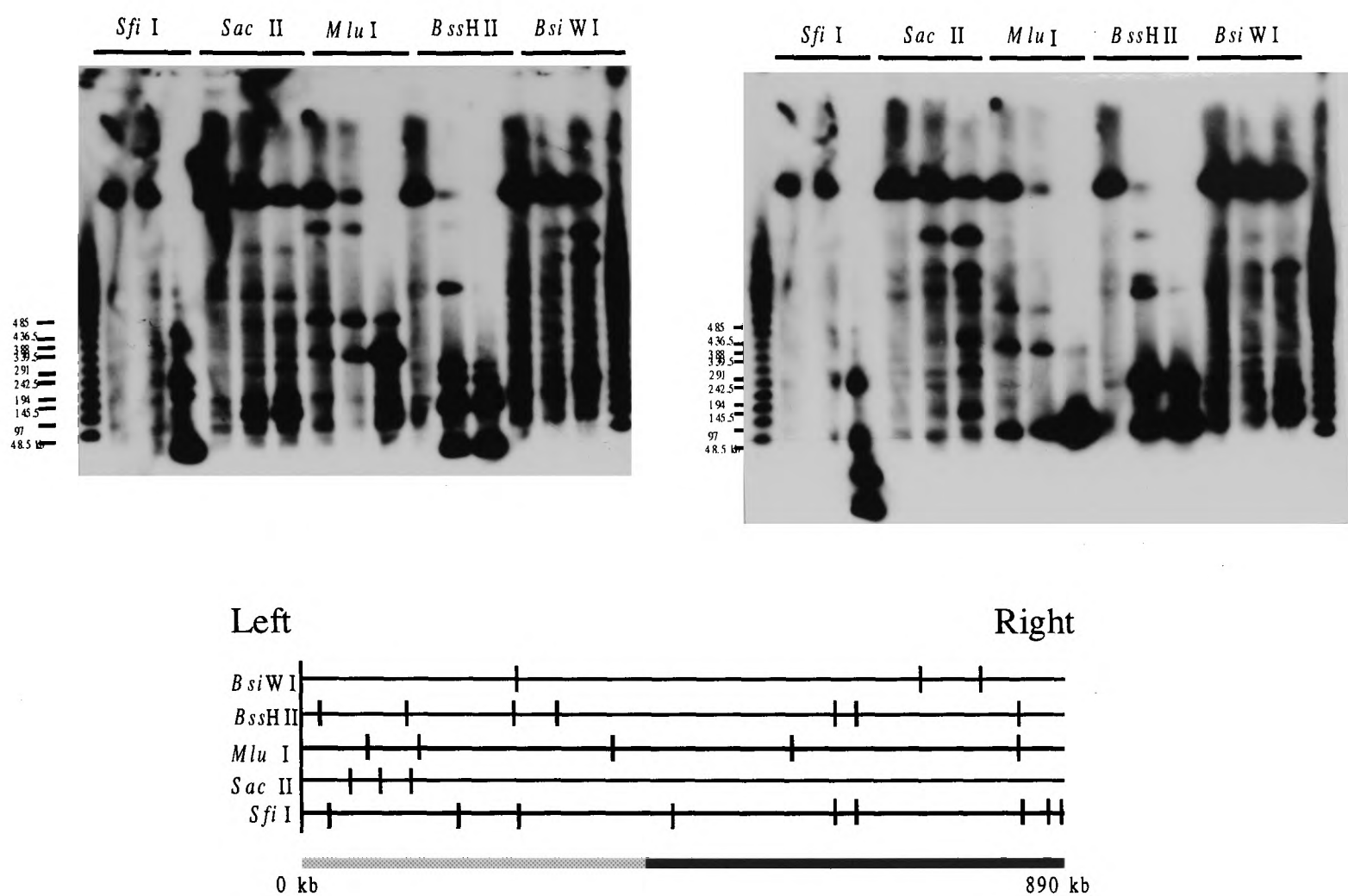


Figure 4.2: Mapping of YAC 692b1. The left-hand side of the page corresponds with the left arm of the YAC, both in the photographs and in the consensus map drawn beneath. The shaded line at the foot of the map marks the chimaeric part of the YAC; the black line shows the Y-chromosomal part.

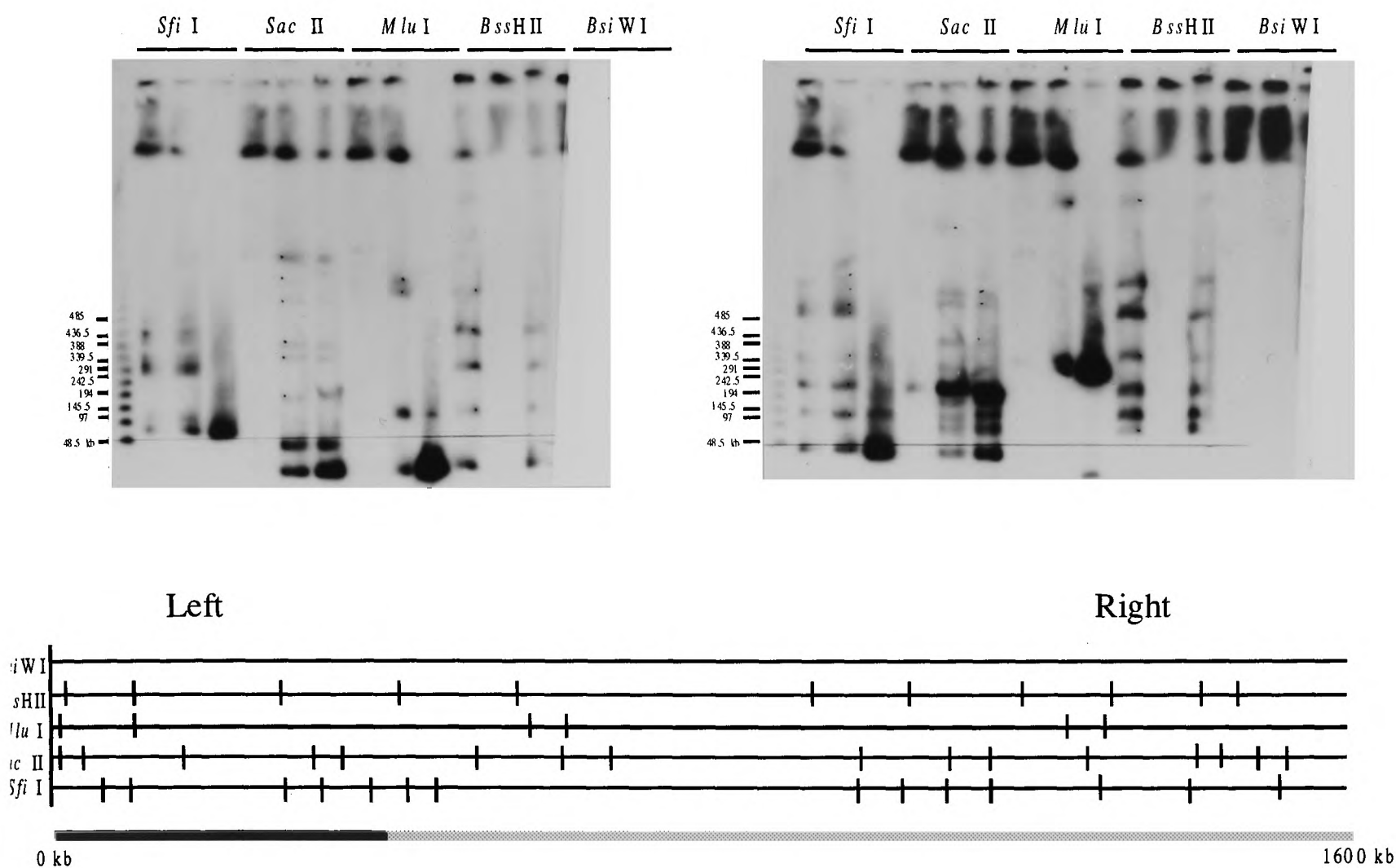


Figure 4.3: Mapping of YAC 872e10. The shaded part of the line beneath the map marks the chimaeric part of the YAC; the black part of the line marks the Y-chromosomal sequences. No *BsiWI* data were obtained.

The left arm of 872e10 and the right arm of 692b1 both fit well with the previous contig map, as can be seen from the maps. 872e10, however, could be seen to be chimaeric, as only the first 410 kb of the YAC agreed with the consensus map. 692b1, on the other hand, extended at least as far as the probe AA15e, and the YAC contained an additional 580 kb of sequence, potentially extending the contig. In order to confirm this region, as well as to derive a novel probe, the left arm of this YAC was end-rescued as a plasmid (as described in section 2.6). After obtaining a plasmid, the insert DNA was excised from the plasmid vector, and hybridised to a panel of *HindIII*-digested DNA from OXEN (49, XYYYYY), normal male and female cell lines, three rodent cell lines containing a human Y chromosome, and two revertant cell lines which had lost the Y sequences. The hybridisation is shown in figure 4.4.

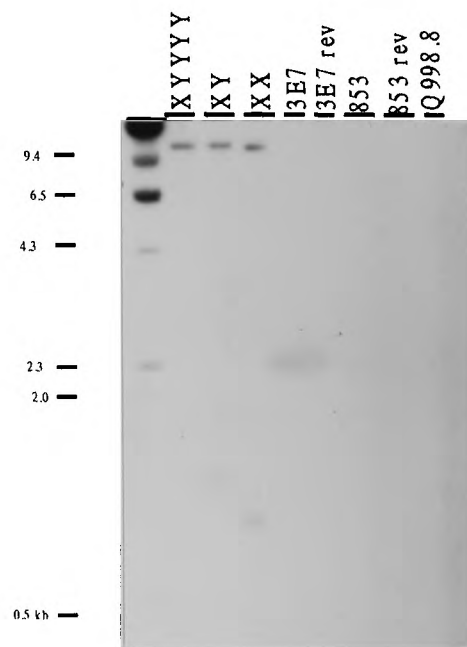


Figure 4.4: Hybridisation of the end-rescued probe 692e to a panel of *HindIII*-digested cell lines containing Y-chromosomal DNA.

As can be seen from the figure, hybridisation takes place in all the human cell lines, but none of the Y-hybrid cell lines. This means that the probe is not derived from the Y chromosome, but is an autosomal sequence. Hence, YAC 692b1 is chimaeric for some part of its length; the safest assumption being the final 400 kb, which is not overlapped by any other YACs at present.

Further mapping data from this region are discussed in section 4.4. Figure 4.5 below shows the locations of the two YACs described above on the contig map.

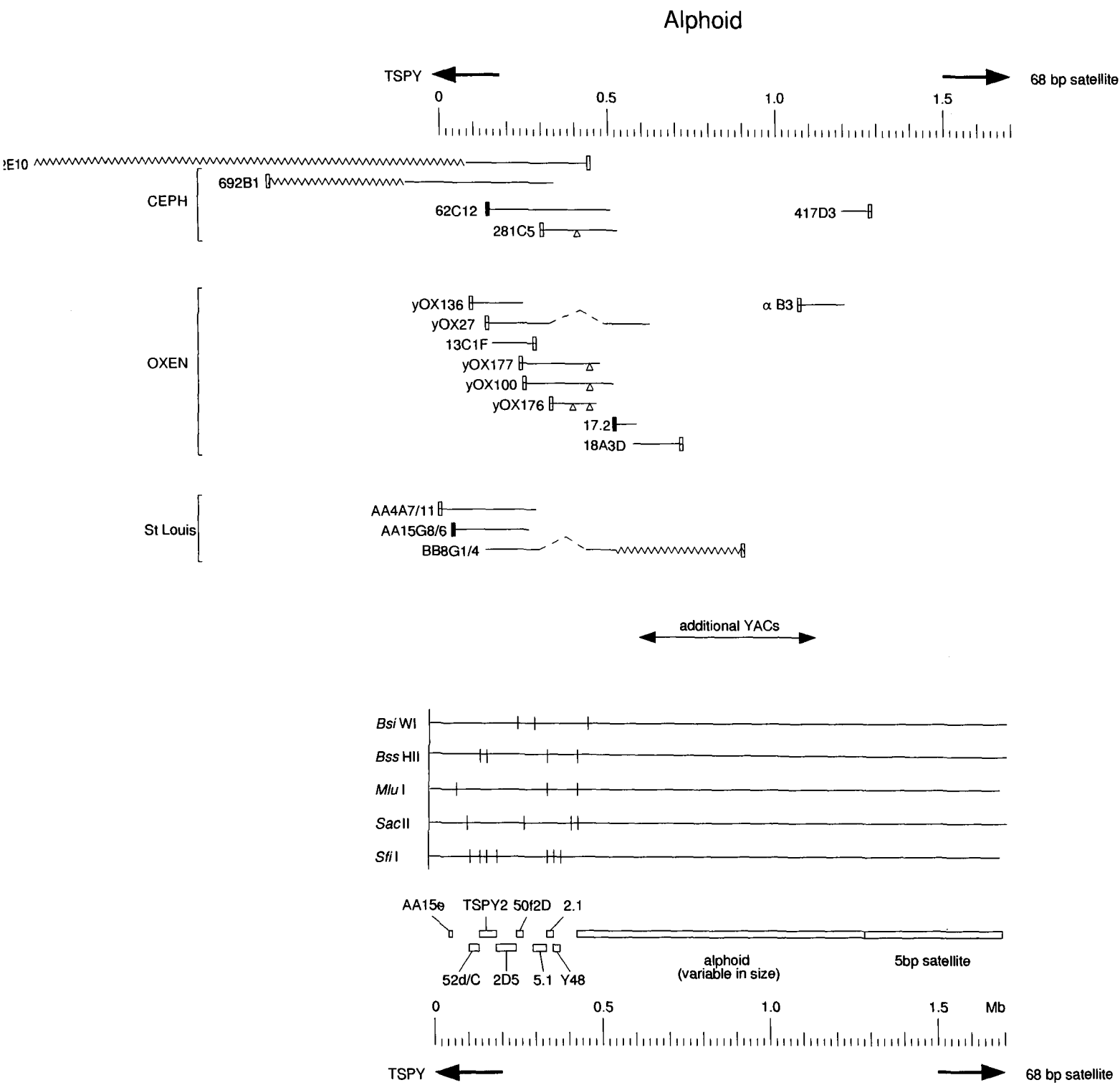


Figure 4.5: Consensus map of the alphoid contig.

4.2 The TSPY contig

This contig is centred on the large block of TSPY1 sequence (also known as TSPY major, or Y-190 major). The TSPY sequence is a block of repeated sequence composed of a 20 kb repeat unit; and is polymorphic in size between individuals. As mentioned in the previous section, there is also a smaller block of TSPY (TSPY2, TSPY minor, Y-190 minor) adjacent to the alphoid satellite block. The two TSPY regions are distinguished not only by the size of the repeated block, but also the flanking sequences – TSPY1 is flanked by 52d/B, rather than the 52d/C found adjacent to TSPY2. It will be recalled from section 4.1 that, in the course of the library screening, the YAC 803d2 was identified as containing TSPY sequences and 52d/B, which places it on the TSPY contig. The mapping of this YAC is shown below (figure 4.6).

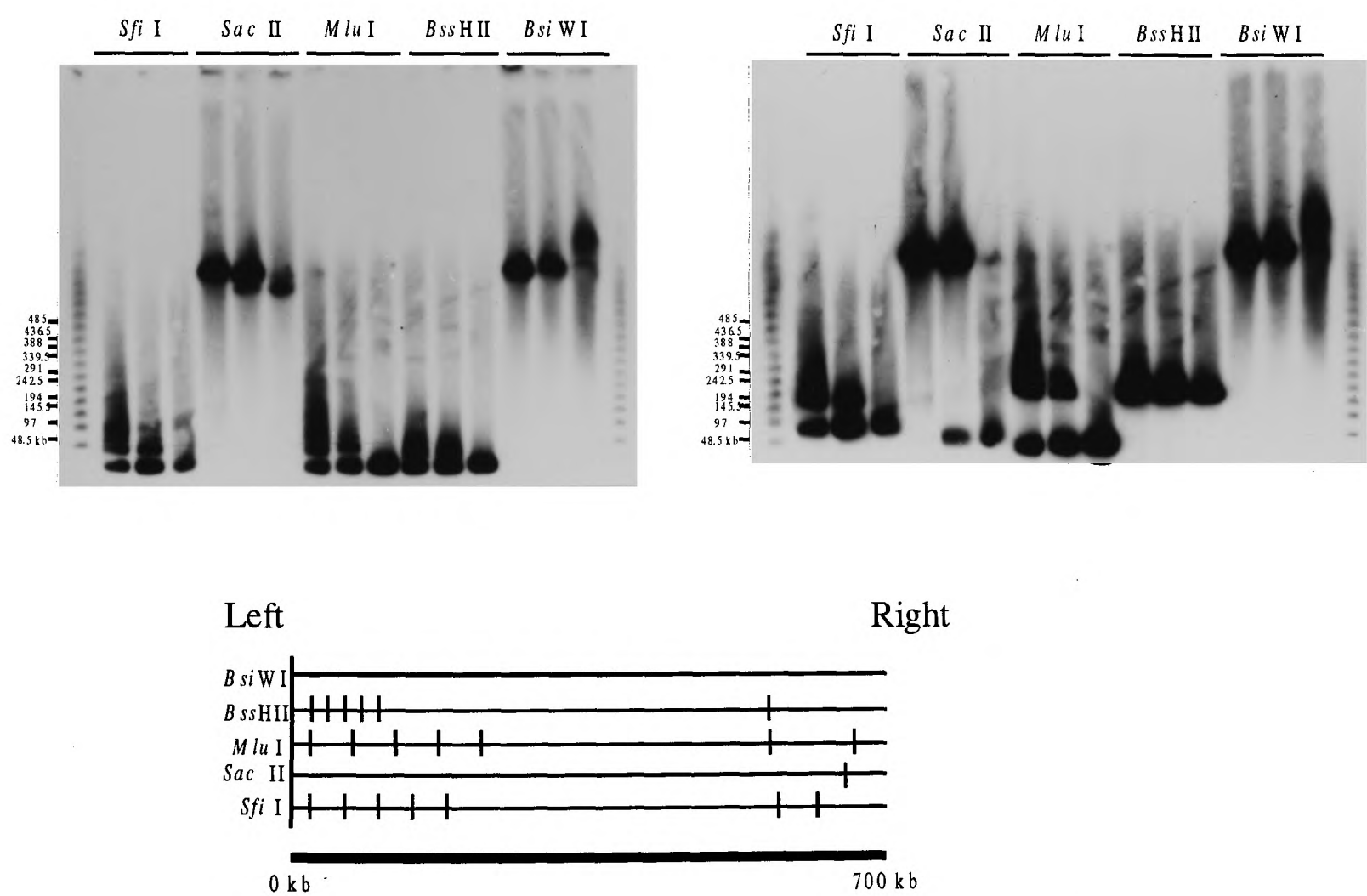


Figure 4.6: Mapping of YAC 803d2. The left-hand side of the page corresponds with the left arm of the YAC, in both the hybridisations and the map underneath.

The YAC fits onto the contig map without any problems – the left end of the YAC contains the TSPY1 sequences, and the right end of the YAC covers the region containing 52d/B. Not all of the characteristic restriction sites of the TSPY1 block can be seen on the partial digests of the YAC, although the sites closest to the end of the YAC can be clearly seen. Unfortunately, the YAC was not found to extend the coverage of the contig at all, being completely contained within a previously mapped CEPH YAC, 759g2*. This YAC, known to us as 759g2, does not correspond with the actual 759g2 YAC obtained subsequently from the HGMP-RC. However, it does unambiguously cover the TSPY contig; while the HGMP-RC 759g2 YAC originates from the *AMELY* contig. Accordingly, our version of the YAC will be referred to as 759g2*.

A subsequent screening of the CEPH mega-YAC library with the TSPY probe alone picked out a number of YACs which seemed to hybridise very strongly to the probe, suggesting that they contained the TSPY1 block, rather than the small TSPY2 region. However, subsequent probing with the flanking probes (52d and LLY22g, another moderately-repeated sequence) produced no positive YACs; and mapping of two of these YACs (772a5 and 684e1) showed neither the presence of the characteristic restriction site pattern of the TSPY1 block, nor any noticeable correspondence with the previously-mapped YACs of the contig. Hence, no further mapping of this contig was possible. (However, see also section 4.4)

The current map, with the position of YAC 803d2 marked, is given below.

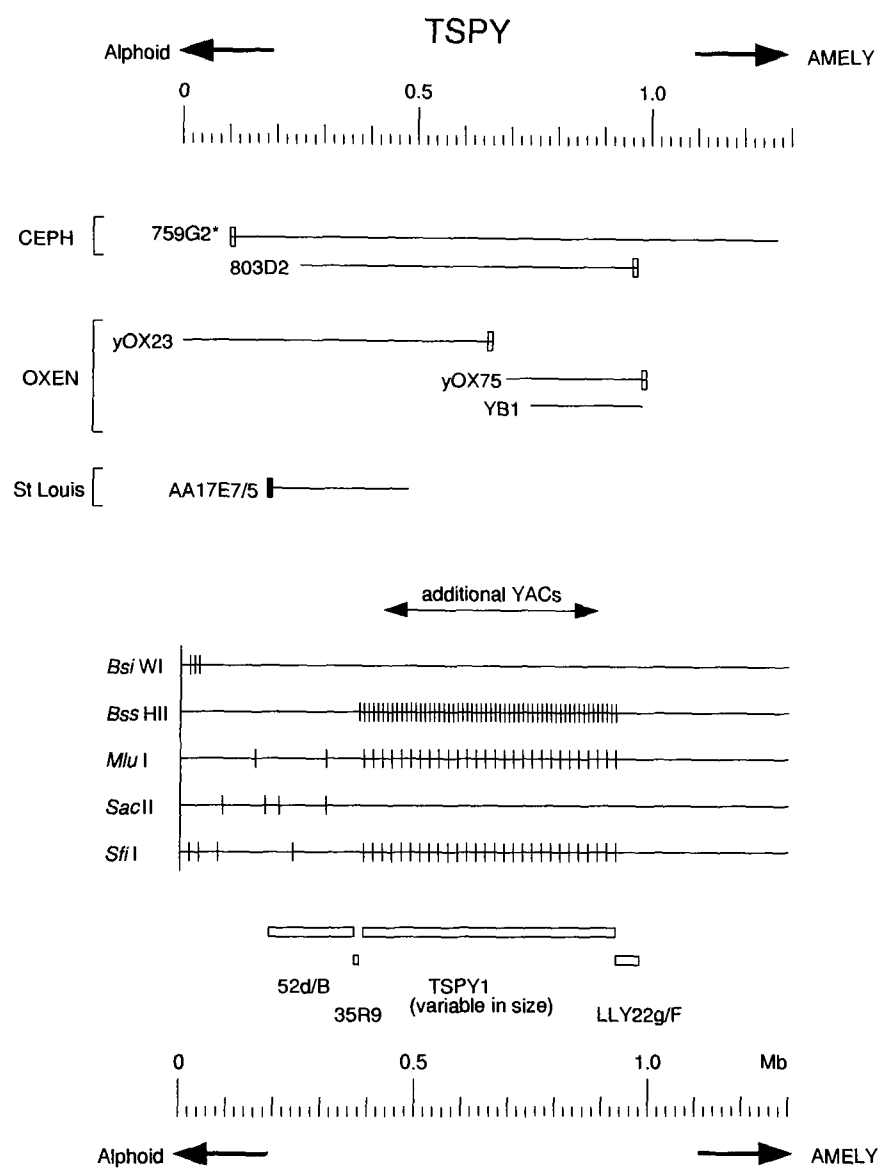


Figure 4.7: Consensus map of the TSPY contig.

4.3 Mapping of the *AMELY* contig

At the outset of this work, the *AMELY* contig was the largest of the maps, covering around 3 Mb; but of this only 1700 kb was covered by multiple YACs, and it was unclear how reliable the remainder of the map was. The contig was assembled around YACs containing the *AMELY* gene, the Y chromosomal copy of the X-Y homologous amelogenin gene. Coverage was densest in this region, consisting of 5 OXEN YACs, 3 St. Louis YACs, and 8 CEPH YACs; two of the CEPH YACs (758g1 and 757h5) providing the unconfirmed 1300 kb of sequence.

It will be recalled from figure 4.1 and table 4.1 that YACs were identified as containing the *AMELY* contig probe AMY8e, in addition to a small block of probes (AA15e, 52d/C, and TSPY) determined as lying on the alphoid contig. It was thought that this may provide a link between the *AMELY* and alphoid contigs, and so YACs 752h1 and 894f6 were mapped. The results are shown below, in figures 4.8 and 4.9.

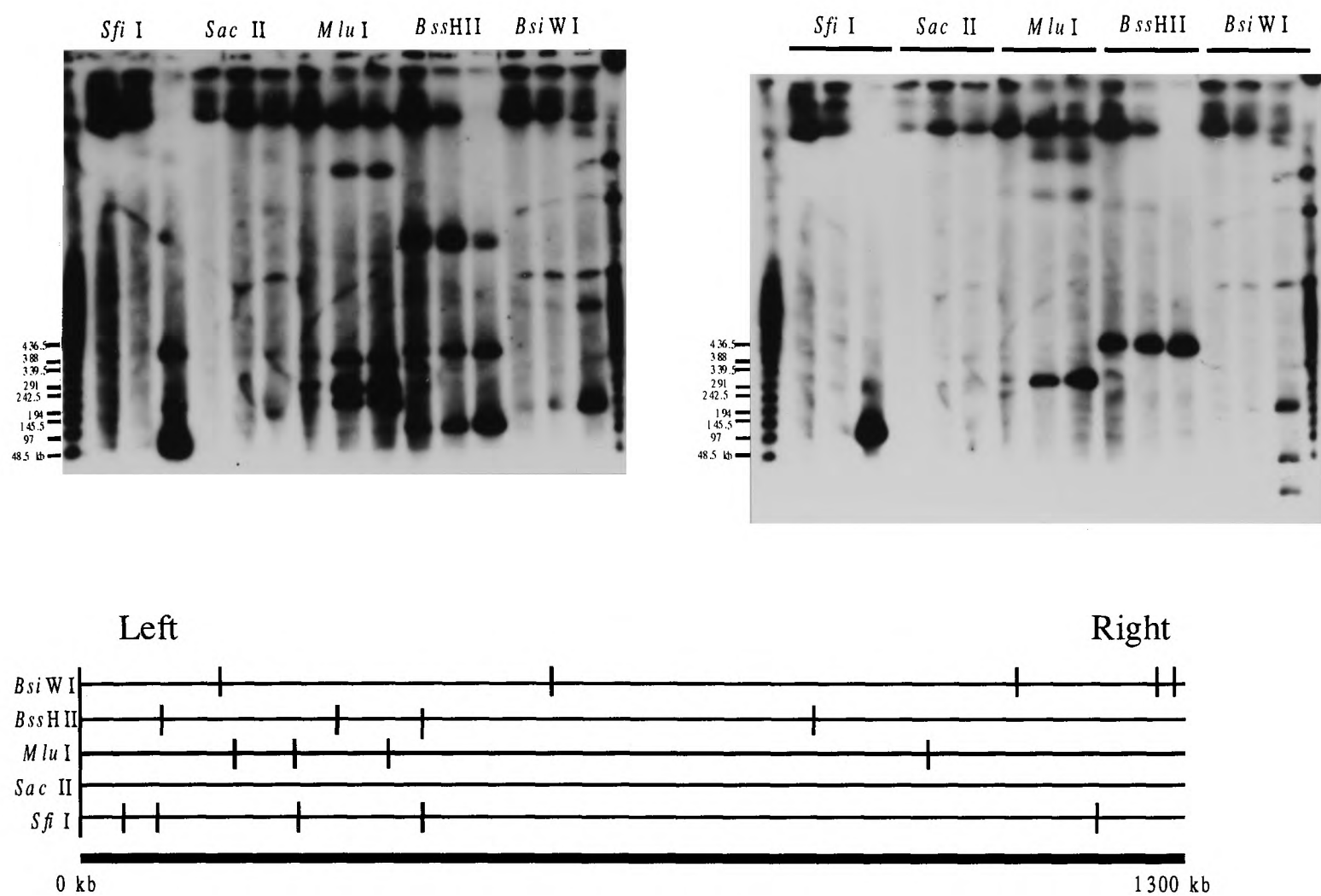


Figure 4.8: Mapping of the YAC 752h1; the left-hand side of the page corresponds with the left arm of the YAC. No *Sac*II data were obtained.

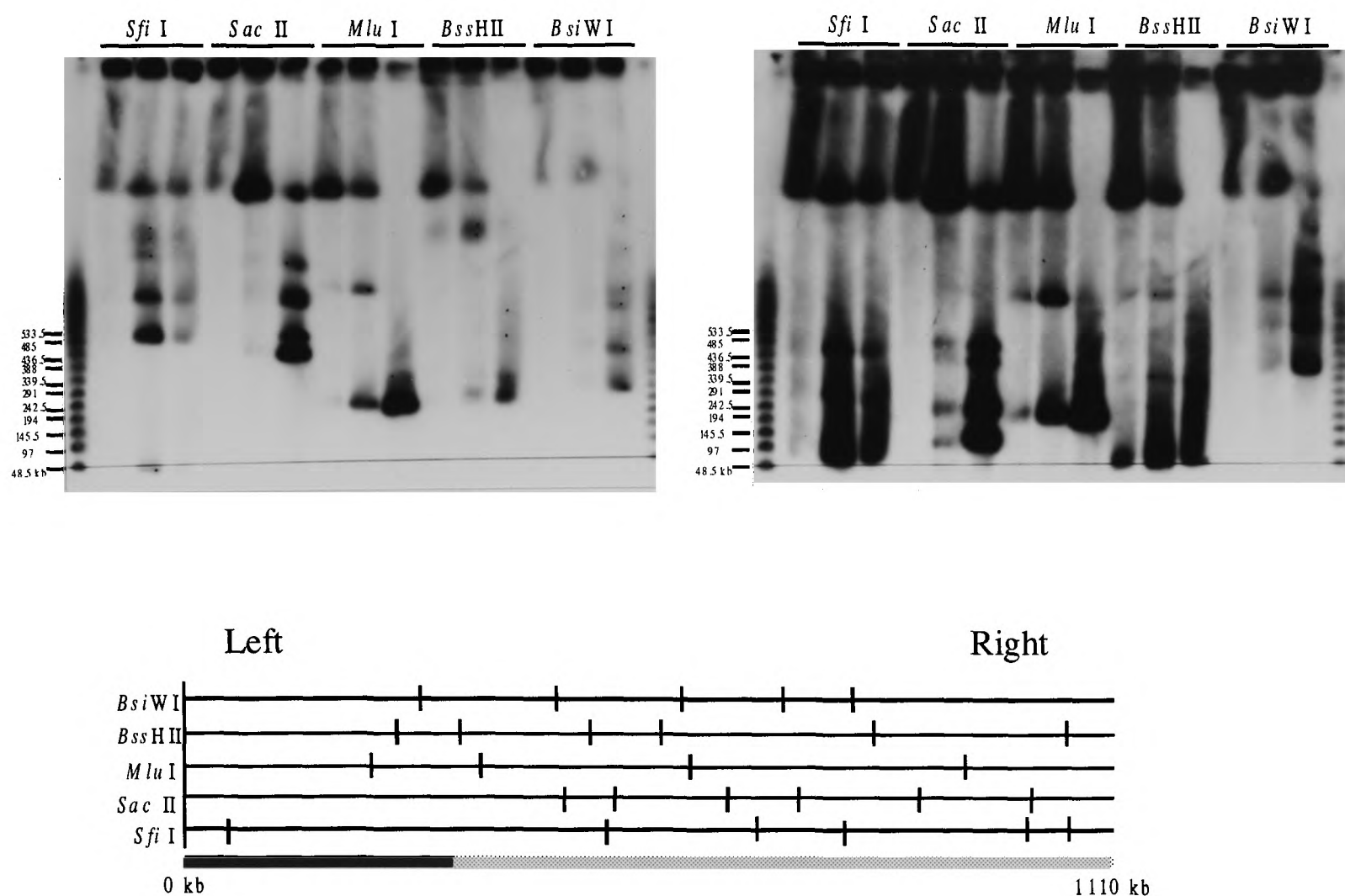


Figure 4.9: Mapping of the YAC 894f6. The left-hand side of the page corresponds to the left arm of the YAC. The line beneath the map shows the chimaeric (shaded) part of the YAC and the Y-chromosomal part (black).

The left arm of 894f6 fitted well with the contig map, in the region containing the probe AMY8e; however only the first 320 kb of the YAC corresponded with the previously-defined map, suggesting that the remainder of the YAC was chimaeric; hence there was no link to the alphoid contig.

YAC 752h1, however, was more interesting. This seemed to fit perfectly with the contig map (the left half of the YAC was contained in the unconfirmed region of the map, overlapping with YAC 757h5). This allowed the map defined by 757h5 to be confirmed, so extending the coverage of the contig. However, 752h1 was also known to contain the alphoid-contig sequences AA15e, 52d/C and TSPY; but no overlap with the map of the alphoid contig was seen. This suggested that there was a duplication of these three sequences on the *AMELY* contig (this being the third copy

of TSPY, it will be referred to as TSPY3). This will be investigated further in the following chapter (section 5.1).

Returning to the physical mapping, as 752h1 extended the coverage of the contig by some 700 kb, it was decided to end-rescue the left arm of the YAC in order to derive a novel probe for further mapping. This was carried out as described in section 2.6; the novel probe 752e thereby created was hybridised to the same digest panel referred to in section 4.1 (for 692e). The hybridisation is shown in figure 4.10 below.

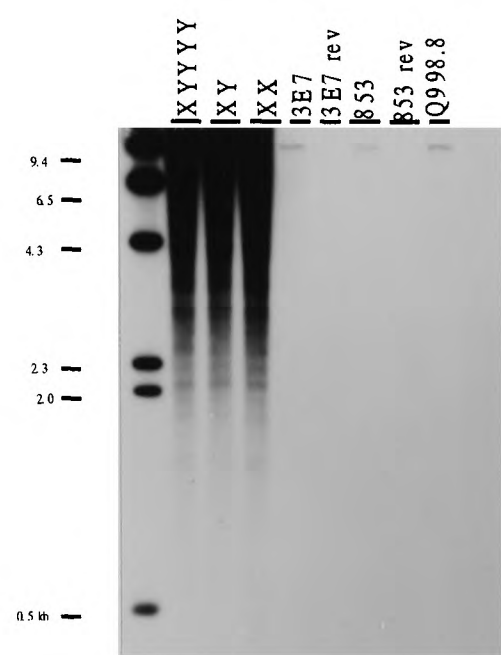
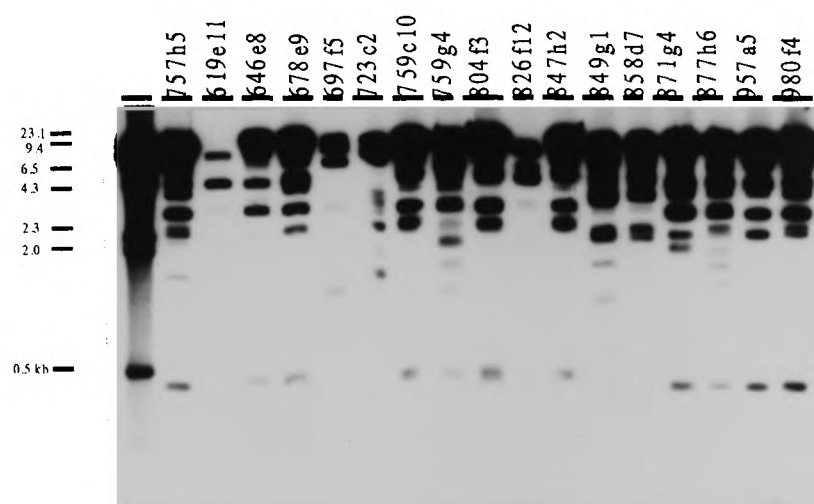


Figure 4.10: Hybridisation of the end-rescued probe 752e to a panel of *HindIII*-digested DNA from human, rodent Y-hybrid and rodent revertant cell lines.

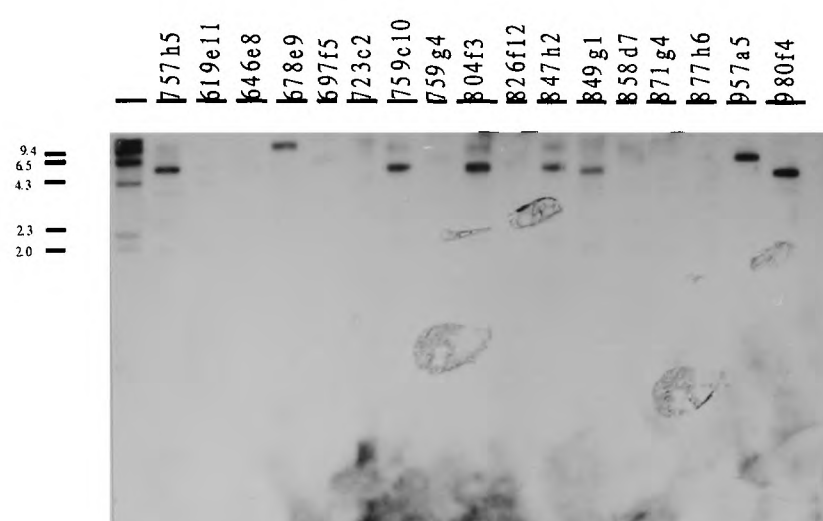
As can be seen from the figure, the probe detects repeated sequences in all human cell lines (including the 46, XX female), but only a single band can be seen on the Y-hybrid cell lines. This suggested that the 752e probe, while detecting a Y-chromosomal sequence, also detected some X chromosomal or autosomal sequences. Nonetheless, it was decided this probe would still be of use in selecting further YACs to map. YAC 757h5 also had previously been end-rescued, to give the probe 757e; however, on hybridisation to Y-hybrid DNA, the probe detected a smear, even with competitor DNA, rather than the single band detected by 752e. Thus, this probe was of no use in screening the library; although it was hoped that it might be of utility in secondary screening of YACs selected by the 752e library screening.

The 752e probe was thus used to screen the CEPH library. Sixteen YACs were selected from the initial library screen; *Hind*III-digested DNA from these YACs was then hybridised with both 752e and 757e. The hybridisations are shown in figure 4.11 below, while the results are listed in table 4.2

Table 4.2: Results of testing 752e YACs with the probes 752e and 757e		
	752e	757e
619e11	-	-
646e8	+	-
678e9	+	-
697f5	-	-
723c2	+	-
759c10	+	+
759g4	+	-
804f3	+	+
826f12	-	-
847h2	+	+
849g1	+	+
858d7	+	-
871g4	+	-
877h6	+	-
957a5	+	-
980f4	+	+



a) 752e



b) 757e

Figure 4.11: The YACs selected with probe 752e were digested with *Hind*III and probed with a) 752e and b) 757e. The results are listed in table 4.2.

Four of these YACs were seen to contain both 752e and 757e (759c10, 804f3, 847h2, and 849g1). These were all mapped, and the maps can be seen in figures 4.12 – 4.15 below. 980f4 also appeared to contain both probes, although the 752e banding pattern differed from the other four YACs, and so this YAC was not studied further.

All four of the YACs covered the same region, and only 847h2 was chimaeric, with 65 kb of non-Y sequence. 849g1 may be chimaeric, as the right arm does not correspond with the consensus map derived from the other YACs. However, the YAC is known to contain both 757e and 752e sequences, which, based on the size of the YAC, does not appear to permit chimaerism. It may be that the rightmost part of the YAC is chimaeric, and by chance contains homologous sequences to the 752e region.

As this part of the map was well covered in other YACs, it was decided not to investigate this anomaly further. Taken together the YACs extended the coverage of the contig by some 1000 kb. Further work carried out with these YACs involved determining the nature of the sequences contained in the YACs (remembering that the probe 752e was either X-Y homologous or had autosomal copies elsewhere in the genome; and 757e was a moderately repeated sequence). This work is discussed in section 5.2.

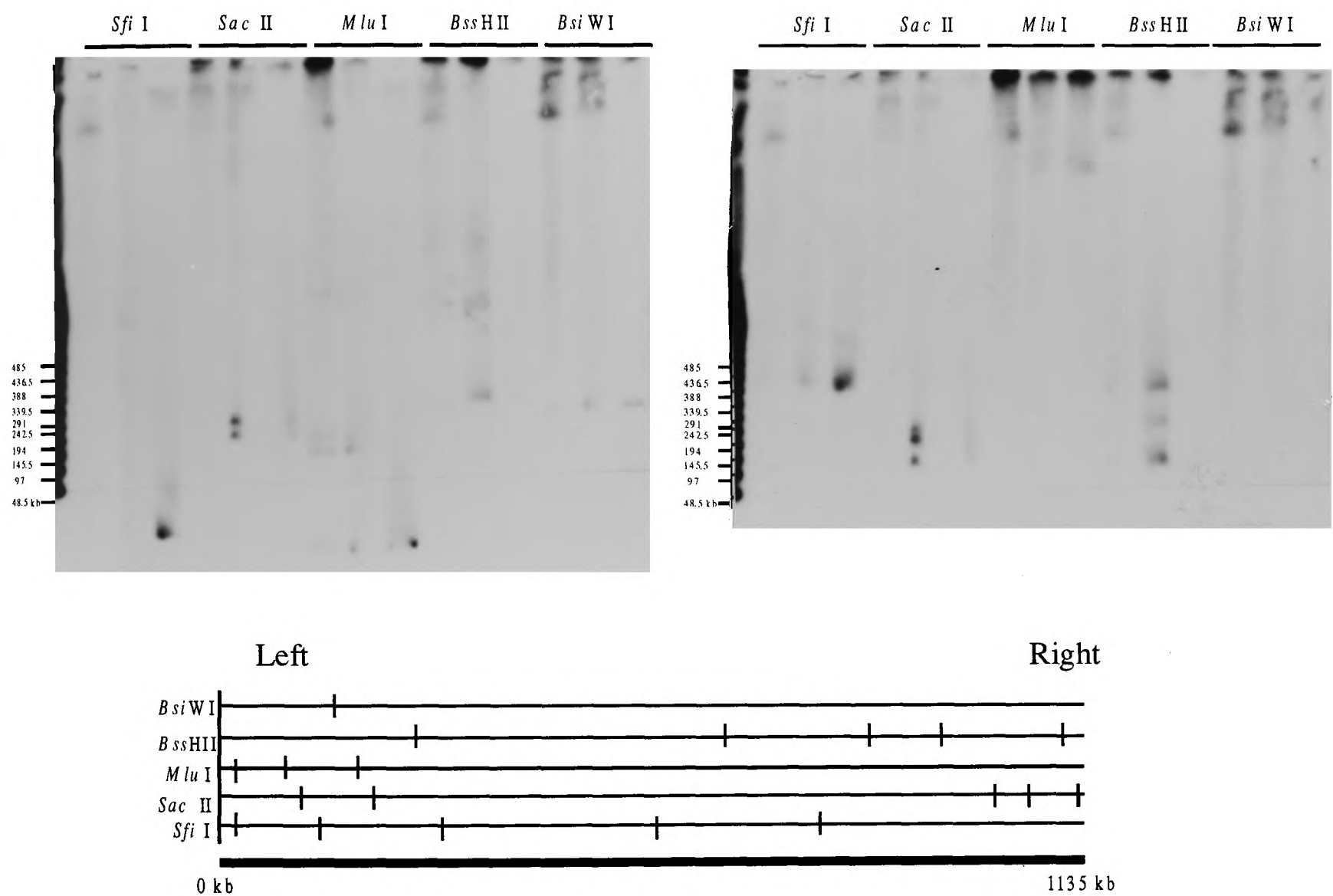


Figure 4.12: Mapping of the YAC 759c10. The left-hand side of the page relates to the left arm of the YAC.

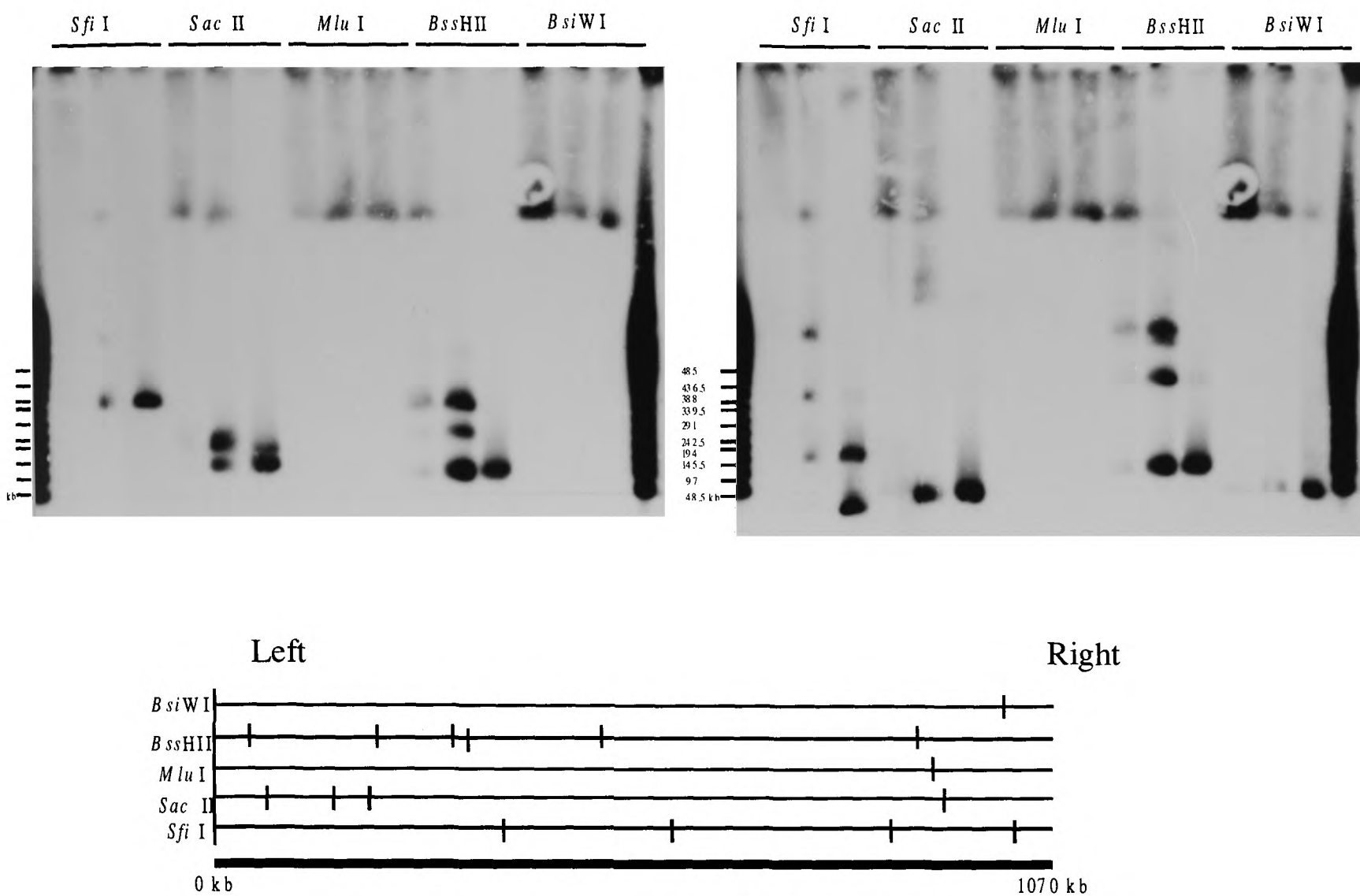


Figure 4.13: Mapping of the YAC 804f3. The left side of the diagram refers to the left arm of the YAC.

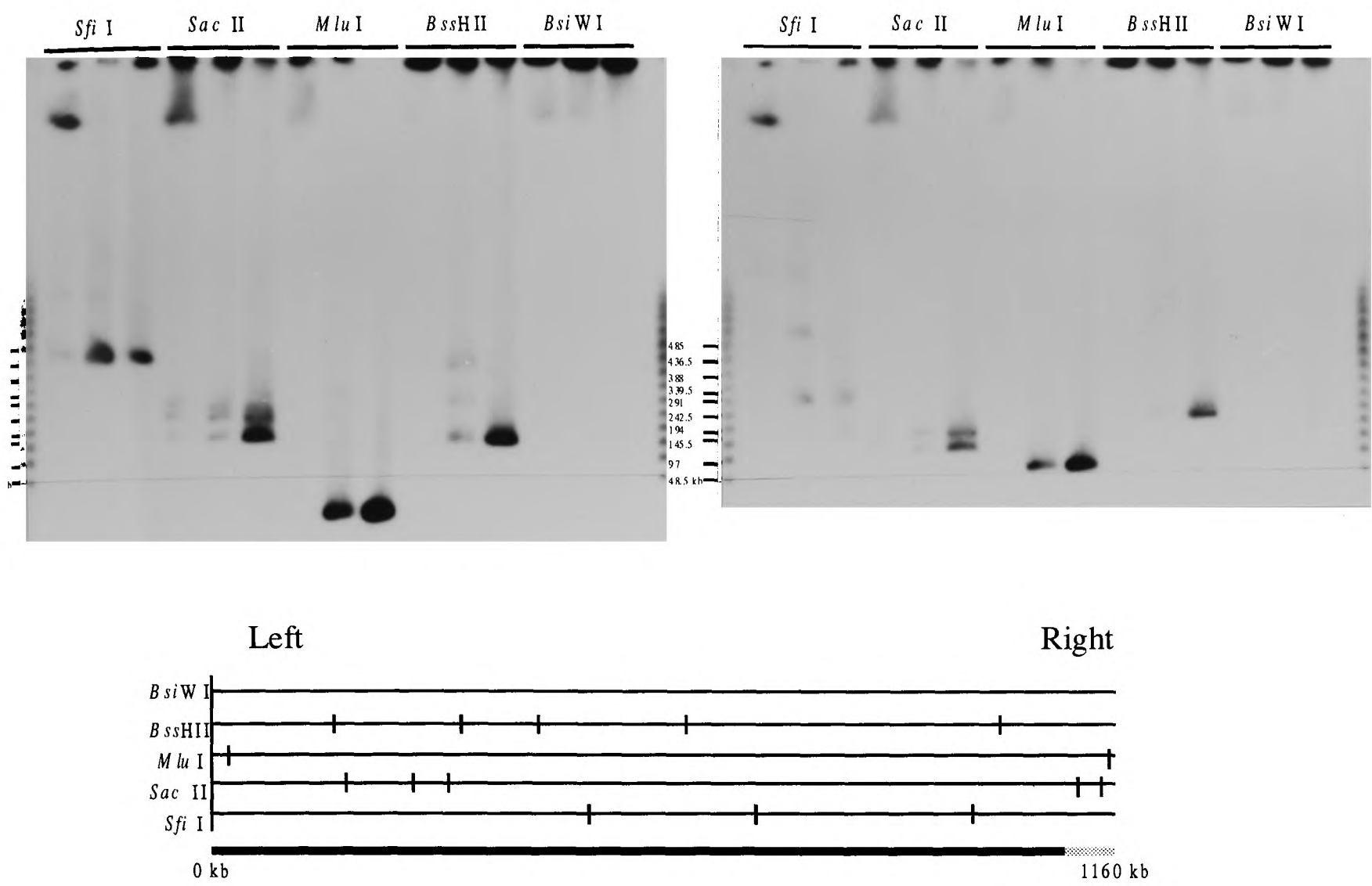


Figure 4.14: Mapping of YAC 847h2. The shaded part of the line beneath the map marks the chimaeric region of the YAC; the black line marks the Y-chromosomal sequences.

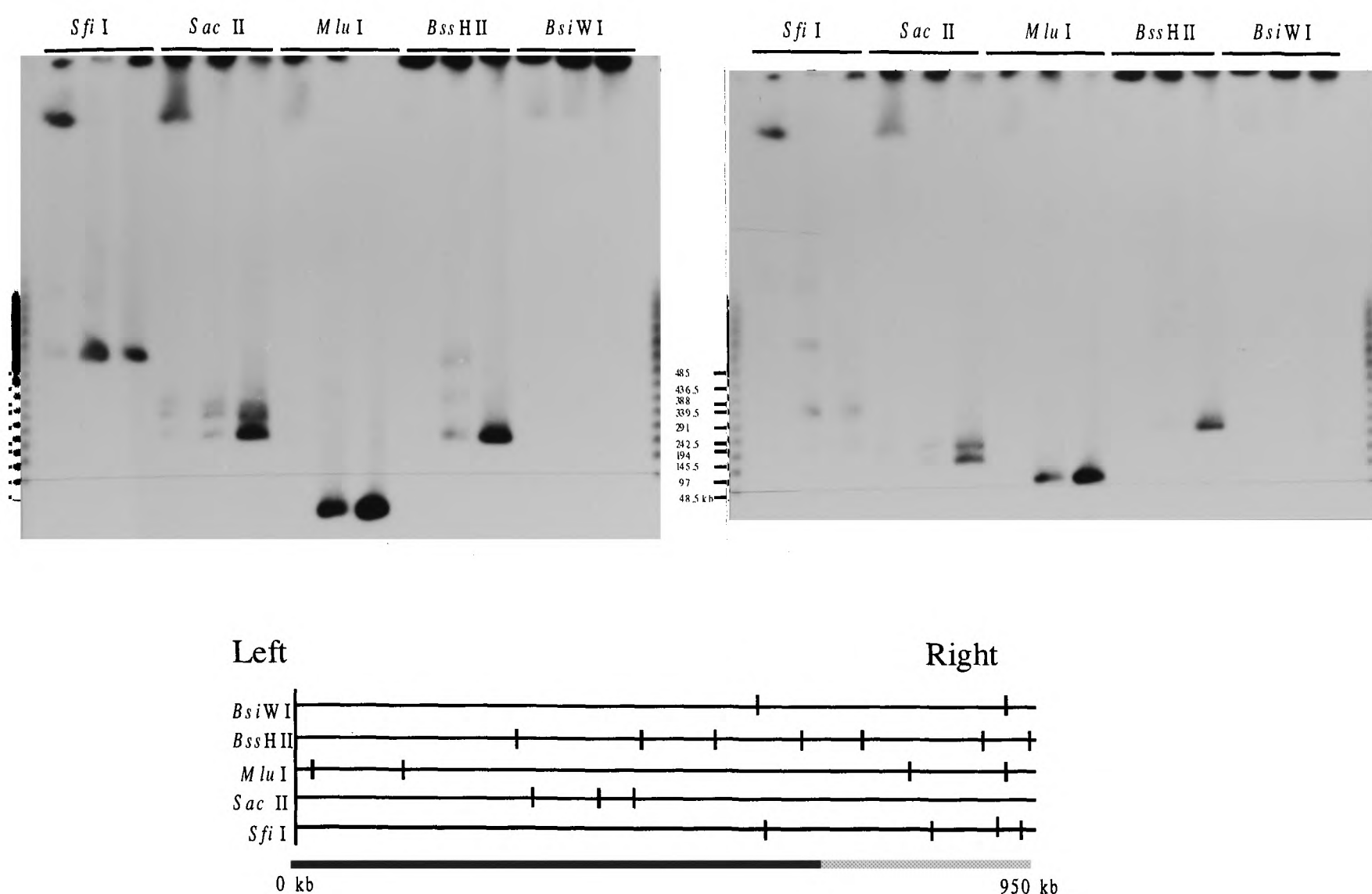
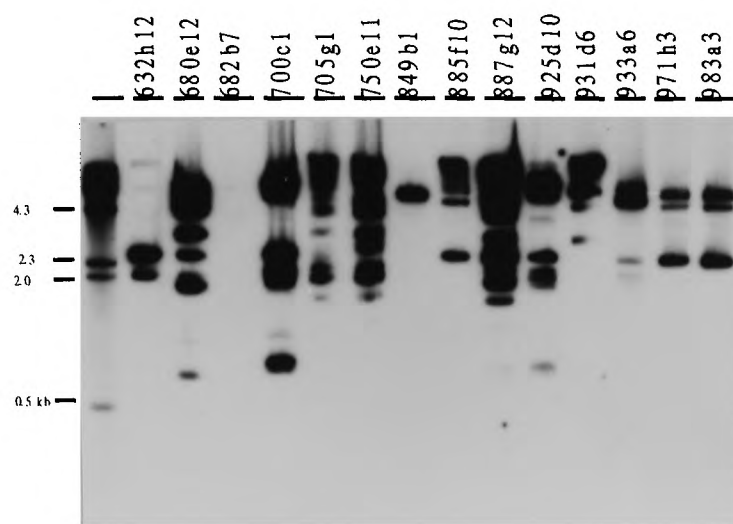
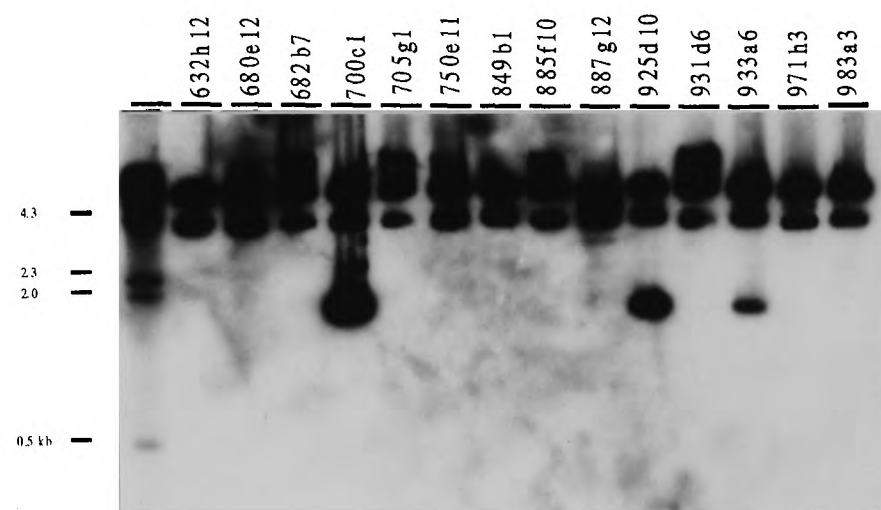


Figure 4.15: Mapping of the YAC 849g1. The left-hand side of the figure relates to the left arm of the YAC. The shaded bar below the map denotes a region of possible chimaerism.

Leaving this side of the contig, YACs were searched for from the other side of the *AMELY* contig. As mentioned above, the most extreme 700 kb of the contig were covered in only a single YAC, 758g1. A probe had been end-rescued from this YAC, although it appeared to give multiple bands on hybridisation to digested genomic DNA. Nonetheless, it was decided to use this probe to screen the CEPH mega-YAC library. It was also known that 758g1 contained the probes 50f2 A and B, so giving a confirmation that any YACs selected were from this region. The hybridisations are shown below in figure 4.16, and table 4.3.



a) 758e



b) 50f2

Figure 4.16: Probing of *Eco*RI digests of the 758e YACs with a) 758e, and b) 50f2. The bands common to all lanes in b) are the result of vector hybridisation to the probe.

Table 4.3: Probe content of 758e YACs		
	758e	50f2
632h12	+	-
680e12	+	-
682b7	-	A or B
700c1	+	E
705g1	+	A or B
750e11	+	-
849b1	-	-
885f10	+	A or B
887g12	+	D
925d10	+	E
931d6	+	A or B
933a6	+	E
971h3	+	-
983a3	+	-

These YACs gave a complex picture to interpret. Although the YACs are marked in the table as being positive or negative for 758e, it is not so clear-cut as that. Those YACs marked as positive actually had a number of different types of banding pattern, presumably reflecting differing subsets of the sequences detected by the 758e probe. Not all of these will be Y-specific sequences. 50f2 also gives a complex picture, although more easily interpretable than the 758e probe. There are at least five different copies of the 50f2 sequence on the Y chromosome, designated by the letters A – E. The YACs which had no 50f2 sequences were not pursued further, as they were not localised in any region of the map. Three of the YACs contained 50f2/E, which is known to lie on the long arm of the chromosome, distal to the GMGY6 contig described in chapter 3; thus these YACs were not studied any further. In addition to this, it was assumed that the 758e-like sequences detected on these YACs were also from the distal long arm, and so any YACs with similar banding patterns from the 758e probe were discarded. One remaining YAC, 887g12, contained the 50f2/D copy, known to be located on the TSPY contig. More will be said about this YAC in the following section.

Of the remaining YACs which had both 758e sequences and 50f2 A or B (the two bands were indistinguishable from these results, although those YACs which were chosen to study further were more closely analysed [data not shown]), only 931d6

displayed the same 758e banding pattern as 758g1 itself (data not shown); thus this YAC was selected to map. The YACs 682b7, 705g1, and 885f10, while containing 50f2/A or B, either did not contain 758e sequences (682b7), or showed a different hybridisation pattern to 931d6, which was later found to come from this region of the chromosome. Thus these three YACs were not investigated further.

However, the YAC 931d6 was very difficult to map – initial analysis showed that the clone contained two different YACs, the smaller of which (around 440 kb) contained the 50f2 and 758e sequences. Thus, it was decided to try to purify the clone further, to obtain a clone containing only the YAC of interest. Initially, the yeast was grown without selection for several generations, to allow loss of one or both of the YACs, followed by selection for the presence of a YAC. This did not have the desired result however, as all clones obtained after reapplying the selection still contained two YACs. It was found that loss of one of the YACs would only occur if selection were for only one of the vector arms (selection was placed on the *URA3* gene, which causes the yeast to require uracil for growth). When this was done, a clone was obtained which only contained the smaller YAC of interest (clone 931-ura-6). The reason for this effect was noticed when mapping the 931-ura-6 clone, when it was observed that the partial digests did not hybridise to the right vector arm probe – presumably each YAC in the cell had lost one of the vector arms, so selection for both arms will result in only those cells containing both YACs. The mapping data from this YAC is shown below (obviously, only mapping data from the left arm is available).

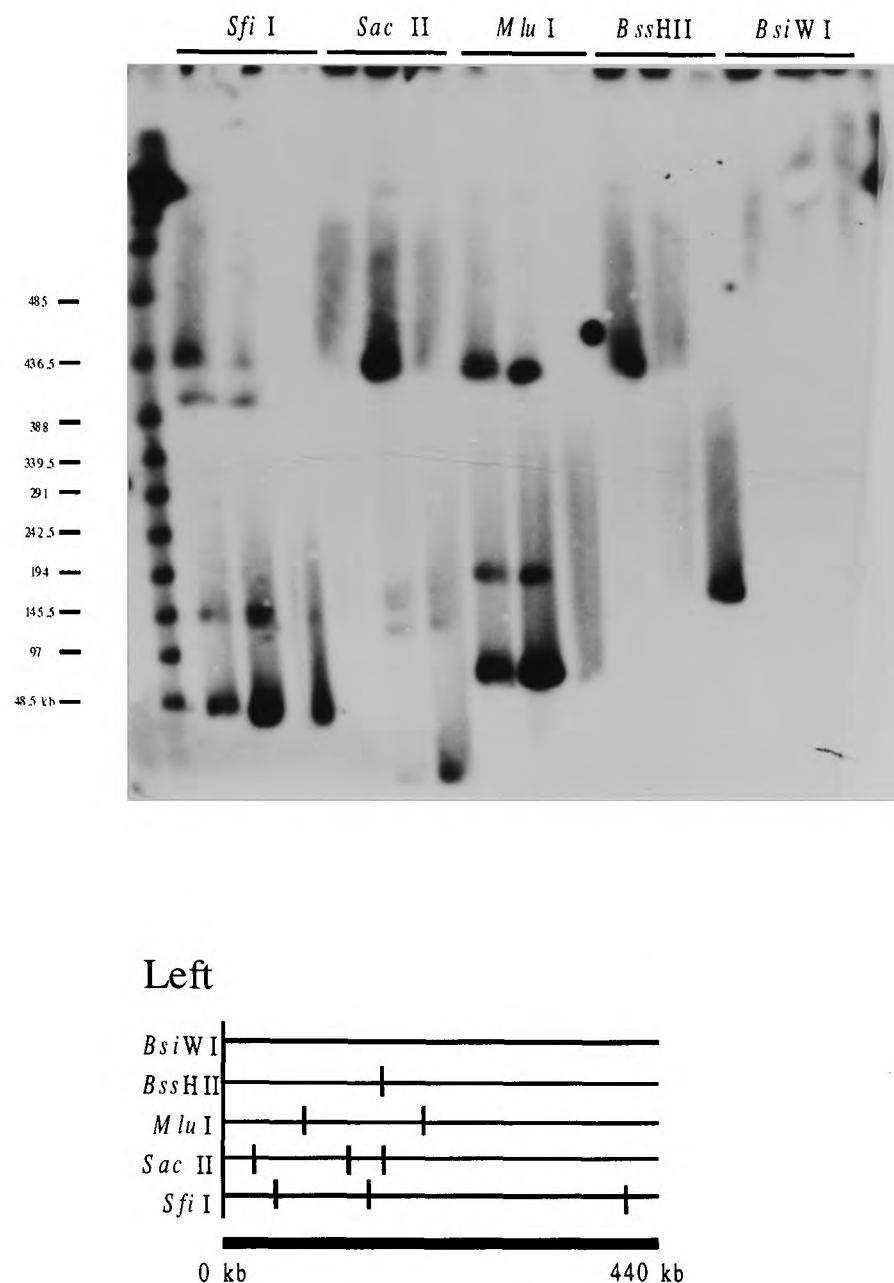


Figure 4.17: Mapping of YAC 931-ura-6. Only mapping data from the left arm is available.

This YAC fits onto the *AMELY* contig map in the expected position, containing both 50f2/A and 758e. Unfortunately, it does not appear to extend the region covered by the map significantly, seeming to end only 20 kb beyond the end of YAC 758g1. Thus, this YAC did not provide any additional probes for screening the library. Additional YACs were isolated which contained 50f2/A, but not 758e, from a screening of the library with the 50f2 probe alone. One of these came from the CEPH library (948a2), and one from the OXEN library (yOX 81). The mapping data from these YACs is shown in figures 4.18 and 4.19 below.

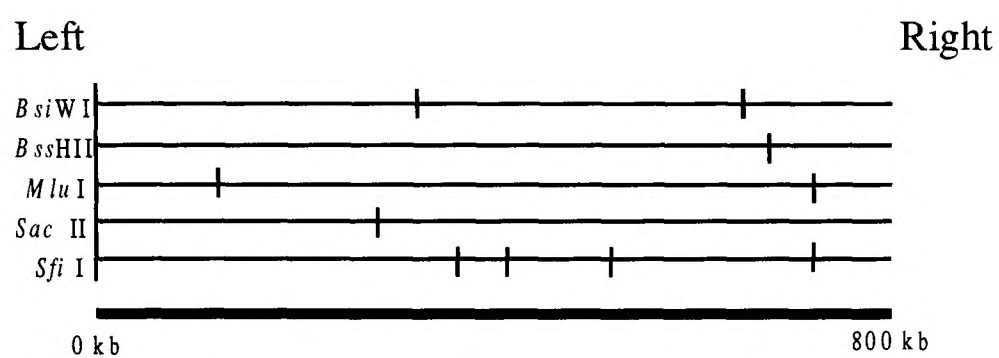
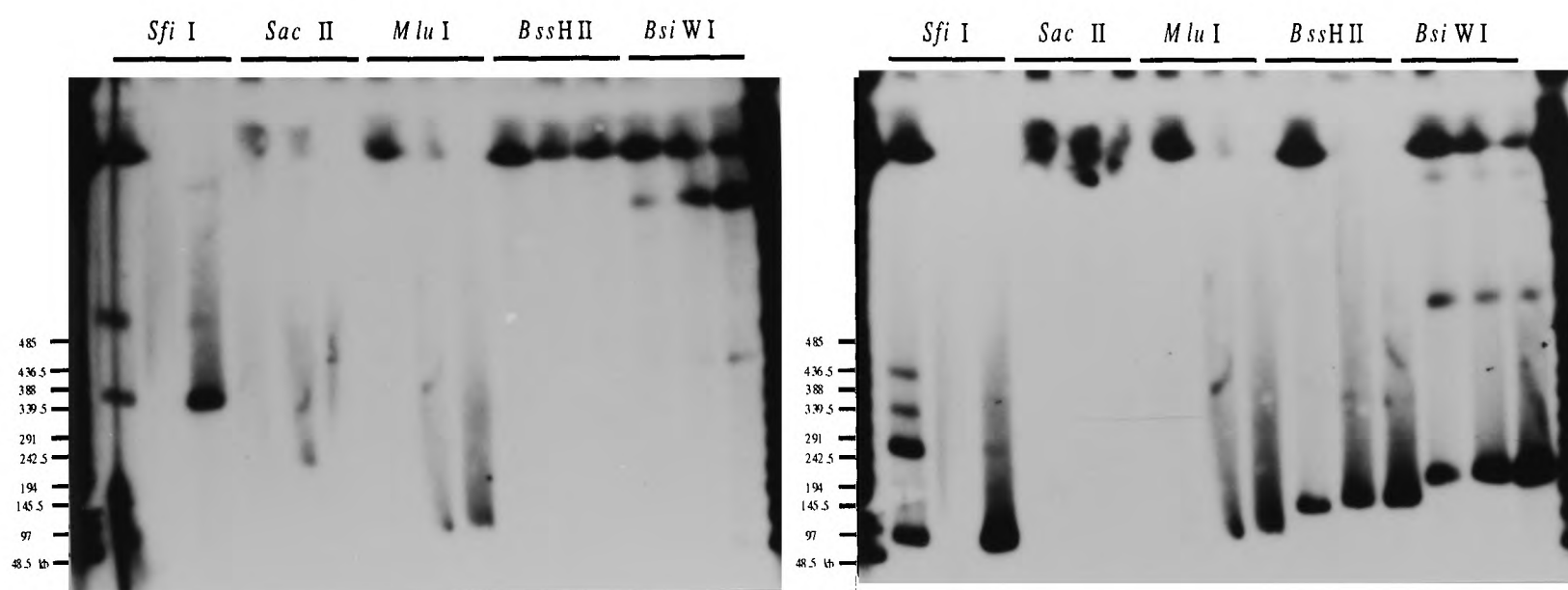


Figure 4.18: Mapping data for the YAC 948a2. The left side of the page relates to the left arm of the YAC.

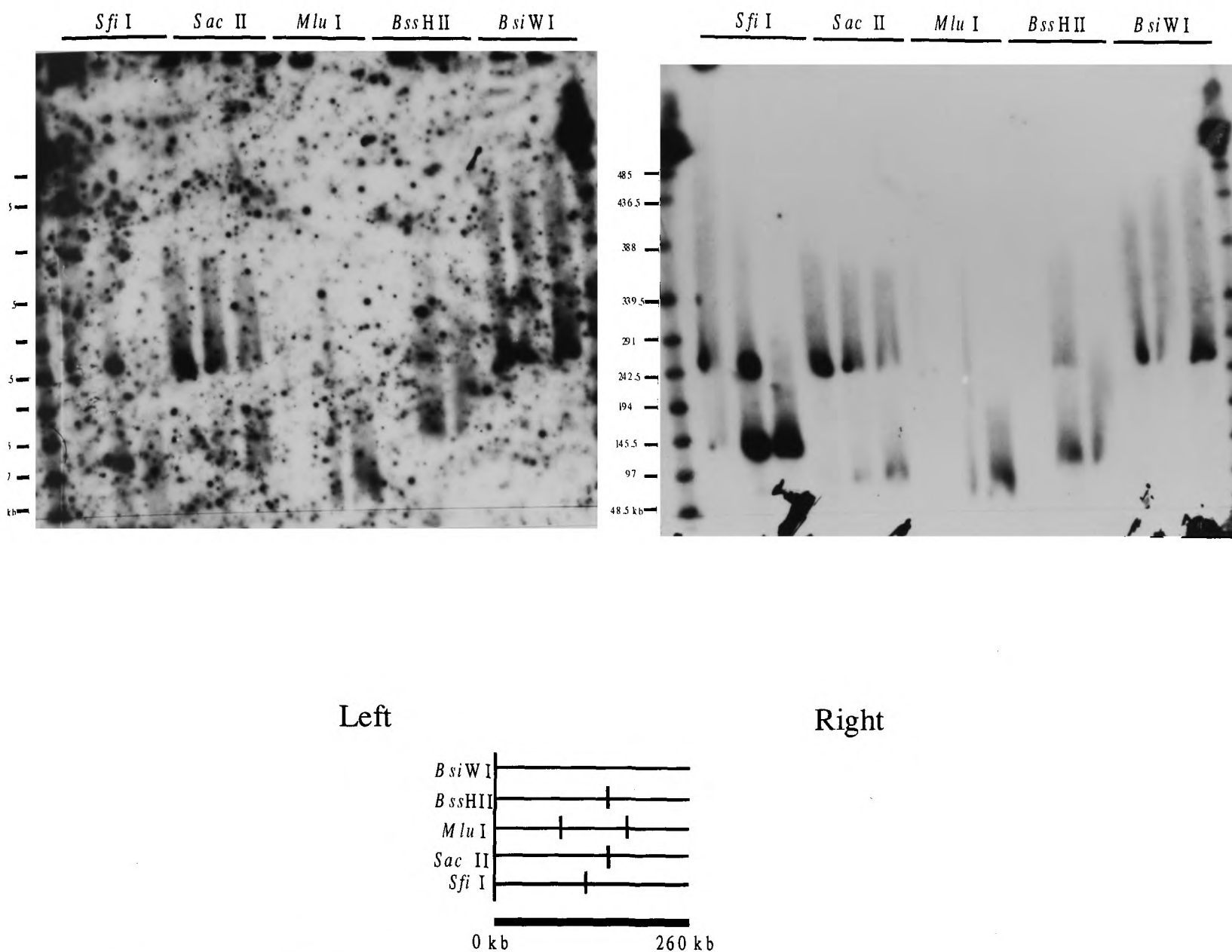


Figure 4.19: Mapping of the YAC yOX 81.

YAC yOX81 mapped to the 50f2 region of the *AMELY* contig, but did not extend the coverage of the contig, although it does provide further confirmation of the map in the region. The YAC 948a2 did not fit with the contig map at all, despite being known to contain the 50f2/A locus. Presumably the YAC is extensively co-cloned, only containing a small fragment of the Y chromosomal region. No firm determination of the chimaeric parts of the YAC was possible. The summary of the contig map at this point is given in figure 4.20 below.

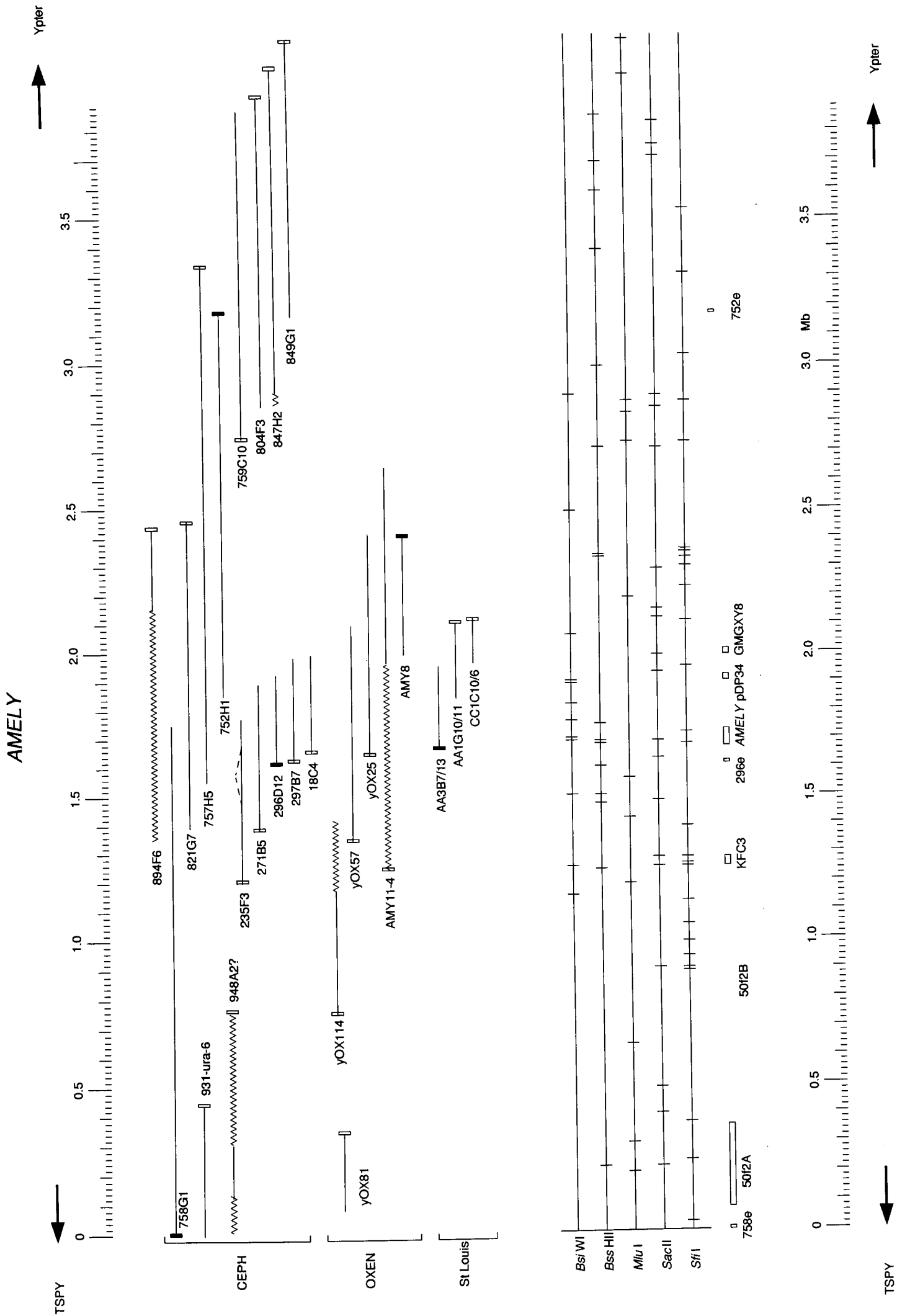


Figure 4.20: Consensus map of the *AMELY* contig.

4.4: The YAC 887g12

This YAC is given a section to itself not only as it was a particularly interesting YAC, but also as it was a particularly troublesome YAC to study and map. Referring back to table 4.3, it can be seen that the YAC 887g12 was originally selected from screening the CEPH library with the 758e probe. However, the secondary analysis of these selected clones revealed that, while 887g12 did have some sequences which were detected by the 758e probe, these did not appear to be the version of 758e which was found on the *AMELY* contig. (It will be recalled that 758e hybridised to a wide range of similar sequences, making it a difficult probe to use.) When the 758e YACs were probed with 52d, it was noticed that 887g12 contained the 52d/B locus, which lies on the TSPY contig. As the coverage of this contig was poor from the CEPH YAC library, it was decided that this YAC warranted further analysis, if only to deepen the coverage of the TSPY contig.

The YAC was digested with *EcoRI*, along with four control YACs of known probe content, and hybridised with the probes 50f2, 52d, and TSPY. The results are shown below, in figure 4.21, and summarised in table 4.4.

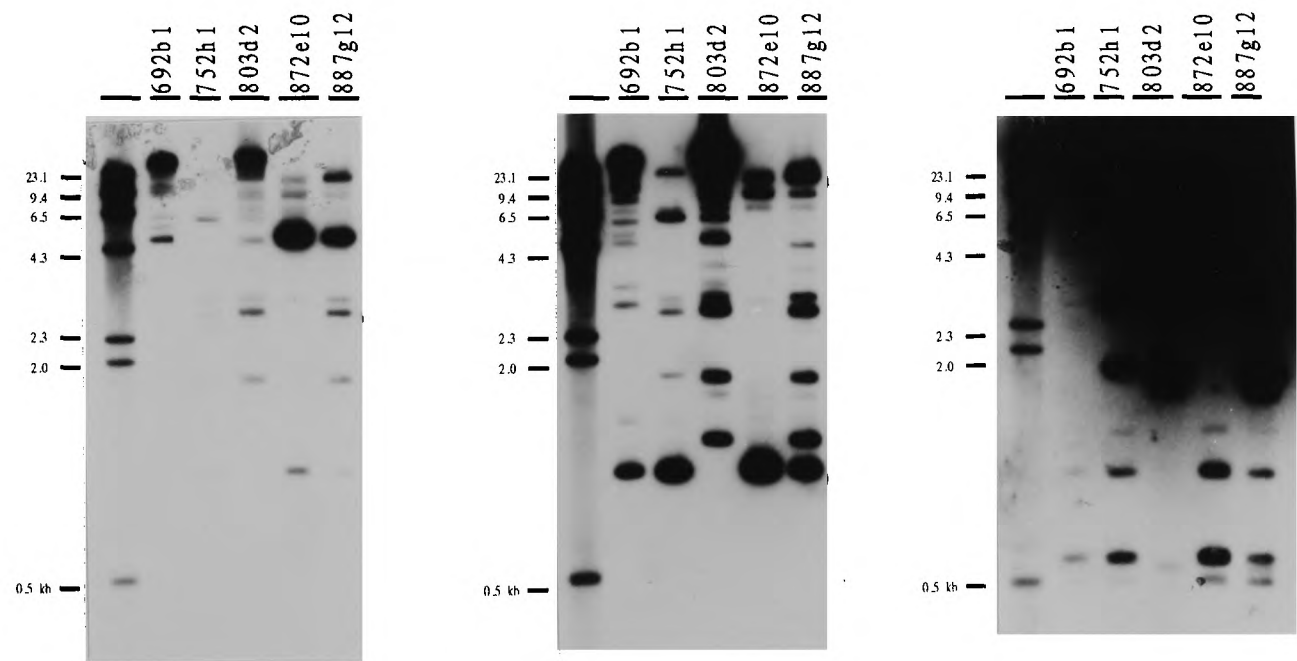


Figure 4.21: Analysis of the YAC 887g12. *EcoRI*-digested YACs were run on a conventional gel, Southern blotted, and hybridised with, from left to right, the probes 50f2, 52d, and TSPY. Several of the digests are incomplete, although the diagnostic 50f2 and 52d bands can be seen.

Table 4.4: Probe content of 887g12, and the control YACs, from figure 4.21			
	50f2	52d	TSPY
692b1	D	C	2
752h1	-	C	3
803d2	-	B	1
872e10	D	C	2
887g12	D	B + C	1 + 2

As can be seen from the table, 887g12 shares sequences from the alphoid contig (50f2D, 52d/C, and TSPY2), and from the TSPY contig (52d/B, and TSPY1). The control YACs enable the three TSPY blocks to be distinguished from their hybridisation patterns. This suggests that 887g12 could provide a link between the alphoid and TSPY contigs; hence it was mapped. The mapping is shown in figure 4.22 below. It will be noted that the 887g12 YAC was not selected when the library was screened with TSPY; some hybridisation was observed, but this was much weaker than the strongest hybridisations, and so the YAC was assumed to contain very little TSPY sequence.

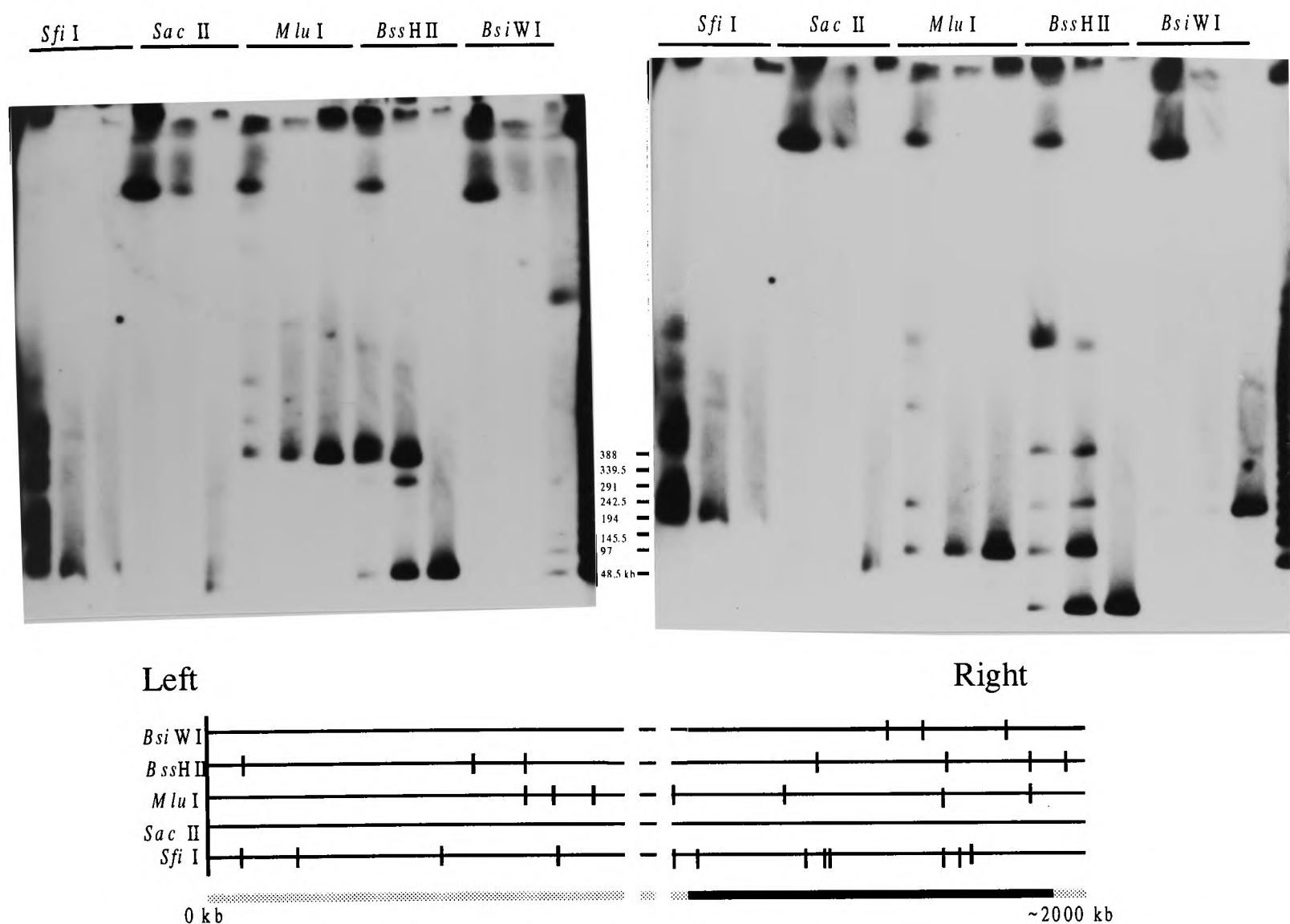


Figure 4.22: Mapping of the YAC 887g12. The left-hand side of the page corresponds to the left arm of the YAC. The line below the map denotes the chimaeric part of the YAC (in grey) and the Y-chromosomal part (in black). The gap in the map marks the unmapped internal part of the YAC. No *SacII* data were obtained.

The map showed that the right arm of the YAC fitted well with the alphoid contig, aside from the endmost 50 kb. However, the left arm did not appear to fit with any part of the TSPY contig, so the YAC must be chimaeric at both ends. A further problem was raised by the size of the YAC, being at least 2000 kb in size. This meant that, at the level of resolution used by the pulsed-field gel, the internal sites of the YAC could not be determined, and so it was impossible to determine if the YAC did link the alphoid and TSPY contigs.

One method of getting around this problem was to try to fragment the YAC, thereby removing the left arm, and reducing it to a size which would be amenable to mapping in this way. The YAC fragmentation vectors pBP108-ADE-CEN and pBP109-ADE-CEN were obtained from J. Heus, Leiden University. These vectors were based upon

the pBP108 and pBP109 vectors described by Pavan et al. (1991), which contained an Alu sequence and a yeast telomere. The vectors used here had been modified by the insertion of *ADE2* and *CEN6* sequences, in order to replace the left arm of the YAC, including the existing centromere. The transformation method used is described in section 2.7.

Upon transformation of the vector, the Alu sequence undergoes homologous recombination with Alu sequences found in the original YAC, and the telomere truncates the YAC, which then loses some of its sequence. Selection for the *ADE2* and *TRP1* genes (*TRP1* is found on the right YAC arm) will select only those transformants which retain the original right arm, and have replaced the original left arm with the transformation vector. Selection against *URA3*, which is found on the original left arm, ensures that only true truncated YACs are identified.

In this way, 9 transformants were isolated, and run out on a pulsed-field gel. This gel is shown in figure 4.23, probed with the right vector arm to show the YACs.

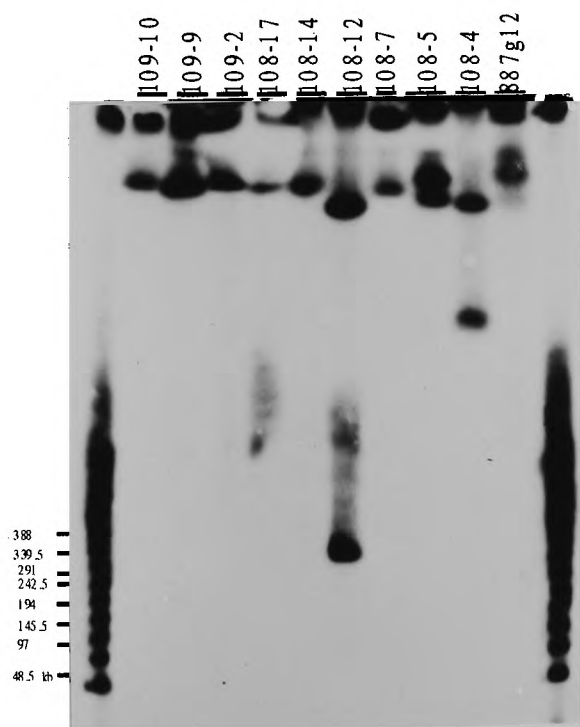


Figure 4.23: Transformants of 887g12 run on a pulsed-field gel, and probed with the right vector arm probe to show the YACs.

It can be seen from the figure that only three of the clones had been significantly reduced in size from the original 887g12 YAC. One puzzling aspect of this was the very low frequency of transformants – only 9 were isolated. Upon inspection of the hybridisation, however, the reason for this became apparent. It will be seen that all of the transformants appear to contain two separate YACs, while the original 887g12 seems to contain only a single YAC. This suggests that the original clone actually contains two copies of the same YAC, which will resolve as only one band on a pulsed-field gel. Thus, any transformants which have truncated only one of the YACs will not survive the round of selection which ensures the original left arm of the YAC has been lost, as they retain one copy of the original vector arm. Only those clones which have truncated both copies of the YAC, or have lost or mutated the second copy, will be selected here, an event which is presumably much rarer than truncating only one copy.

Clones which contain two YACs are not easily amenable to mapping, as the two clones will interfere with one another and obscure the true map. Thus it was necessary to attempt to cause the transformants to lose the larger version of the YAC. The clones 108-4, 108-12, and 108-17 were thus grown for several generations without selection, to allow for loss of one copy of the YAC, and selection was then reimposed in order to retain those clones which still contained a single YAC. Once this was done, the analysis of the deletion clones could begin.

First of all, the clones (hereafter referred to as 887 Δ 4, 887 Δ 12, 887 Δ 17, and 887 Δ 0 [as the original undeleted form of the YAC]) were digested with *Eco*RI, size fractionated on a conventional gel, and analysed for probe content. This is shown in figure 4.24.

It can be seen that both 887 Δ 0 and 887 Δ 4 contain 52d/B and 52d/C, as well as TSPY1 and TSPY2; while 887 Δ 17 contains only 52d/C and TSPY2 (i.e., the alphoid contig loci). A faint band can be seen in 887 Δ 12 in the 52d/B location; as the mapping of this YAC shows it to lie entirely within the alphoid contig, it must be concluded that this result is the effect of a low frequency of clones still being present which contain an additional larger YAC. Thus, 887 Δ 4 was chosen to map as being most likely to close

the gap between contigs, and would be amenable to mapping, as its size was measured as being only 1200 kb, well within the range of mappable YACs. It was also decided to map 887Δ12, in order to obtain possibly more detail of the contig map from the alphoid contig. The maps of these YACs are shown in figures 4.25 (887Δ12) and 4.26 (887Δ4).

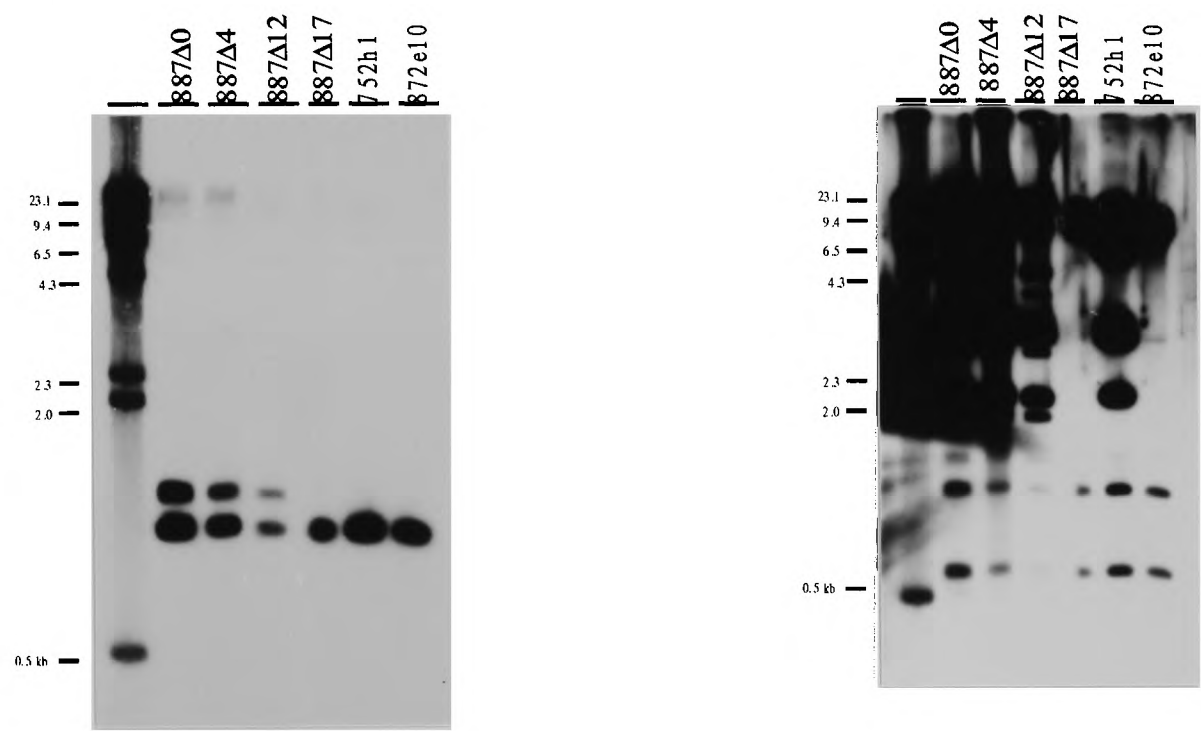


Figure 4.24: Probe content of the 887 deletion YACs. YACs were digested with *Eco*RI, transferred to a nylon membrane by Southern blotting, and probed with 52d (left) and TSPY (right).

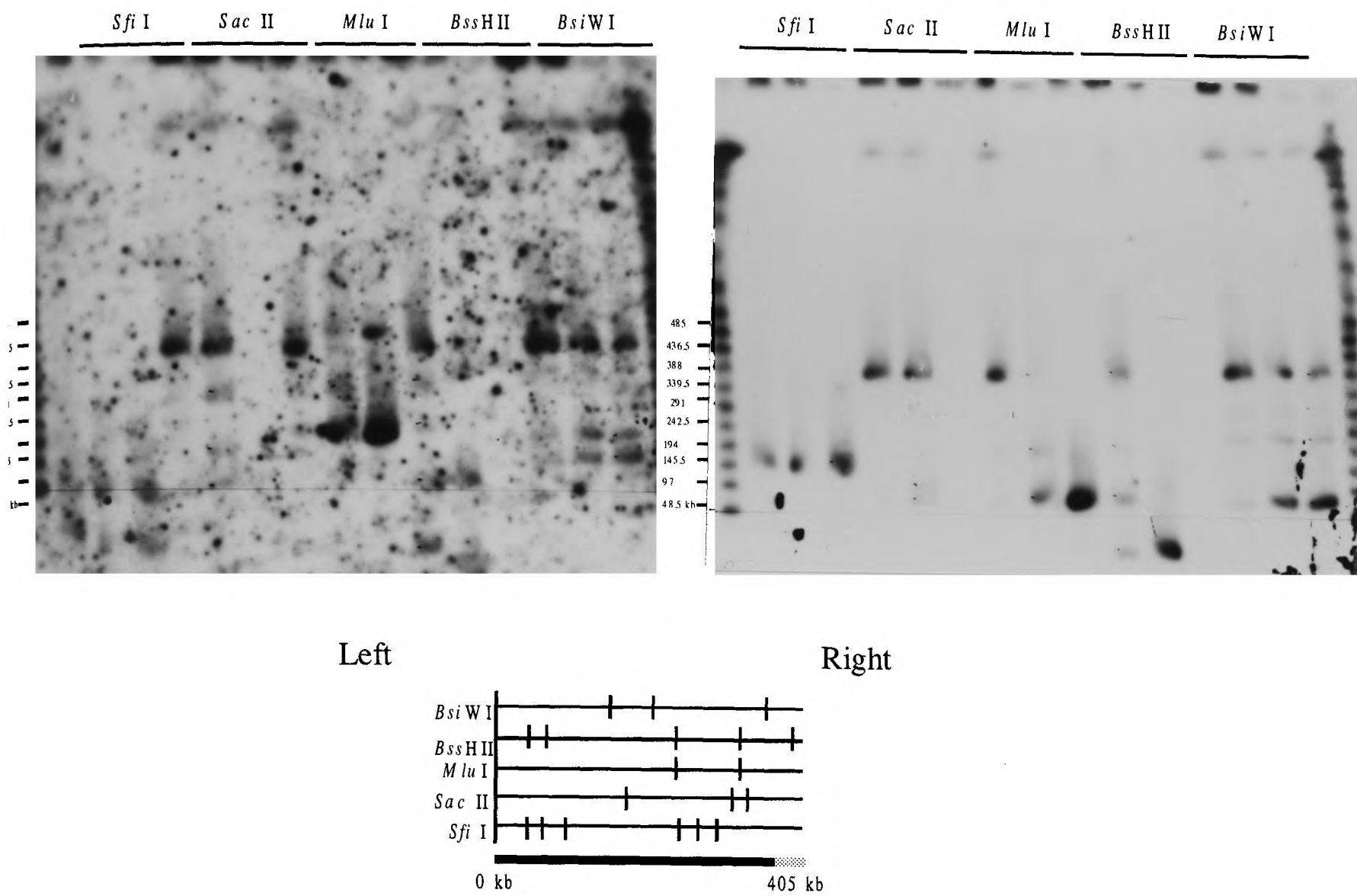


Figure 4.25: Mapping data from the YAC 887-delta-12. The left-hand side of the figure relates to the left arm of the YAC.

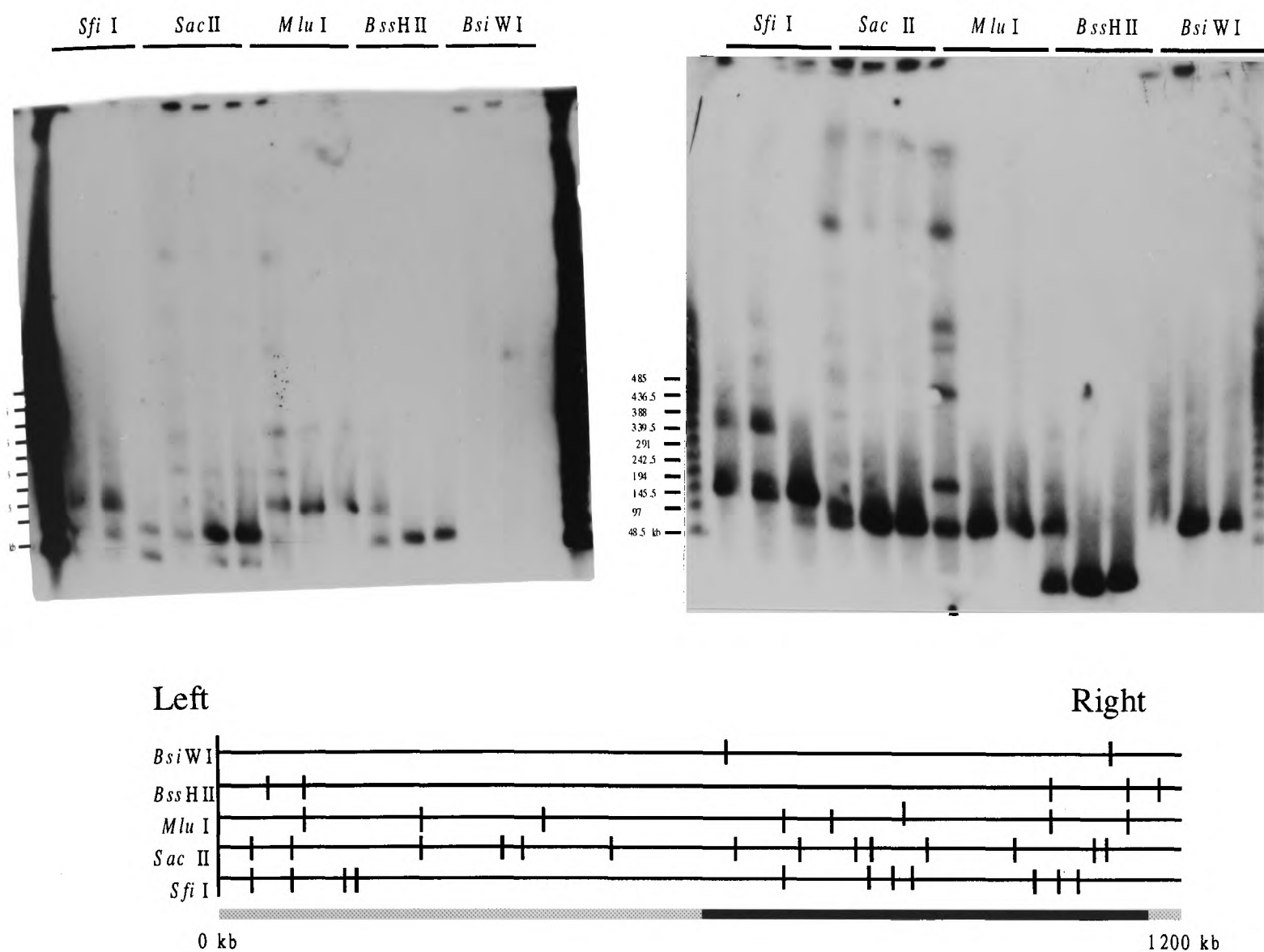


Figure 4.26: Mapping data from the YAC 887-delta-4. The left side of the page relates to the left arm of the YAC. The bar below the map marks the chimaeric region (shaded) and the Y-chromosomal region (black) of the YAC.

While the map from 887 Δ 12 confirmed that this YAC lay entirely on the alphoid contig (further analysis showed that it was truncated within the TSPY2 array), YAC 887 Δ 4 extends from this contig to the TSPY contig. The leftmost 600 kb of the YAC is still chimaeric, but the smaller size of this YAC relative to 887g12 allowed the internal sites to be identified with more precision; in this way it was apparent that this YAC covers part of the 52d/B region of the TSPY contig. The TSPY sequences themselves appear to begin only some 50 kb within the Y-chromosomal part of the YAC, explaining why this YAC was not selected during the TSPY library screen. The revised contig map, linking alphoid to the TSPY contig is shown below in figure 4.27.

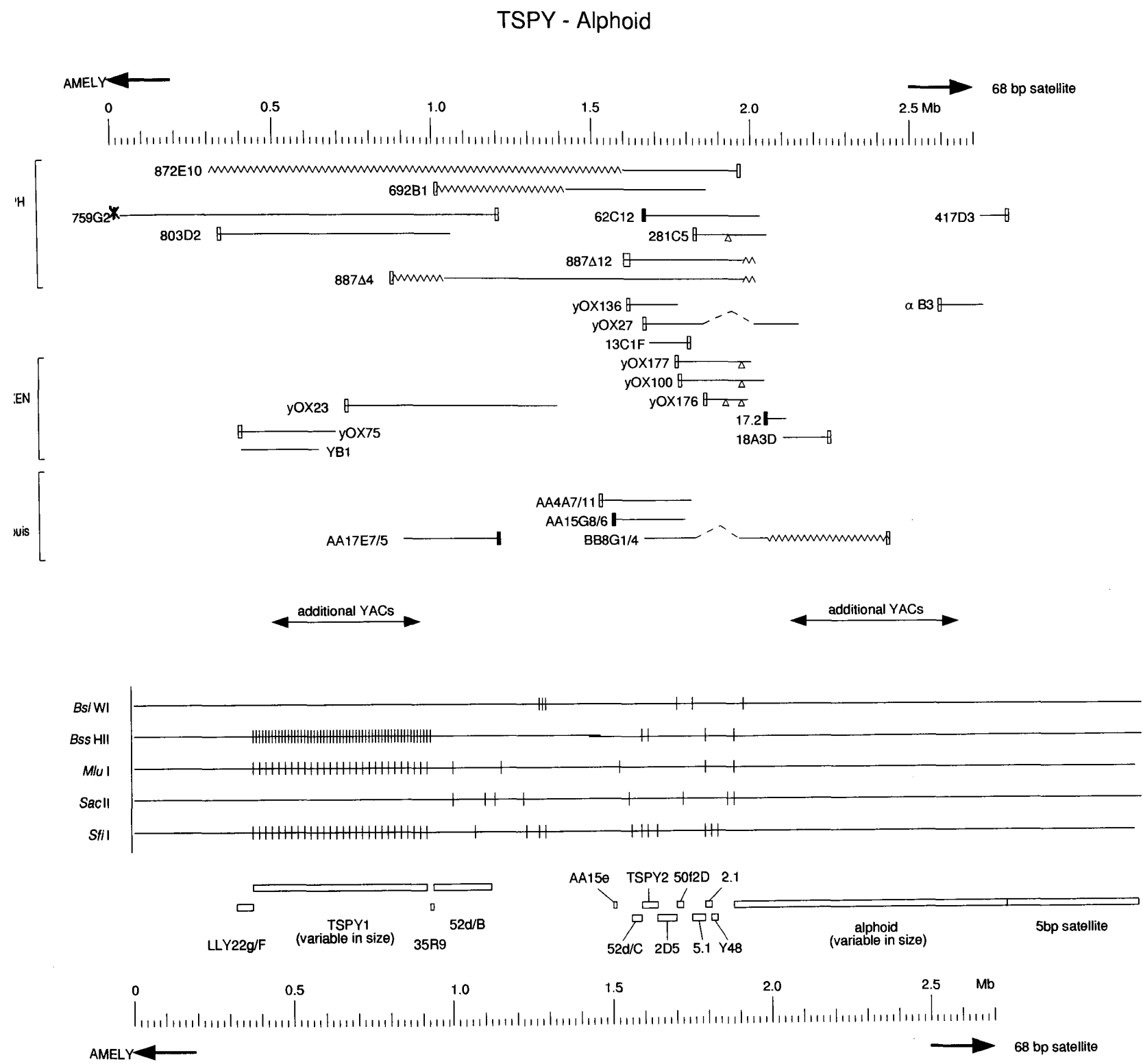


Figure 4.27: The unified TSPY-alphoid contig.

4.5: Closing the remaining gaps

At present the short arm is covered in two contigs, with the one remaining gap being between the TSPY and the *AMELY* contig. The orientation of these contigs with respect to one another is known, and the gap extends from the LLY22g/F side of the TSPY contig to the 758e side of the *AMELY* contig. Despite hunting for YACs to close this gap, none have yet been found. So what are the prospects for closing this gap, and obtaining a single contig covering the short arm of the chromosome?

The main difficulty lies in obtaining suitable probes with which to screen the YAC library – both LLY22g and 758e are repeated sequences, detecting at least 8 loci in the case of LLY22g, and an uncharacterised number of sequences in the case of 758e. This not only hinders the selection of YACs, but also prevents determination of the size of the gap from genomic DNA, as multiple signals would be obtained from both probes. Obviously, further probes need to be obtained in order to close this gap. However, it may be difficult to obtain probes beyond 758e, as none of the YACs so far mapped extend beyond this region, and selecting further YACs to map will be difficult thanks to the repeated nature of the 758e probe.

As for obtaining probes from the TSPY contig, again, this appears to be difficult. The only possible candidate YACs from which to end-rescue probes are yOX75, or 759g2*. Since yOX75 ends in the same region in which the LLY22g probe is found, this would not be likely to provide a useful probe. Similarly, the right arm of 759g2* was amplified by means of Alu-vector PCR; the sequences thus obtained did not appear to be of any use, yielding complex hybridisation patterns even with YACs. And, of course, while another YAC has been isolated which contains the TSPY1 region, it does not appear to map to this contig, suggesting that at best it only contains a small block of TSPY1 sequence in an otherwise chimaeric YAC.

It would appear that the best hope of proceeding to close this remaining gap would be to make use of some other form of cloned DNA library – Y chromosomal cosmid or PAC libraries do exist, and may well contain this missing region, since such libraries

do not suffer from the same instability problems as do YAC libraries. Unfortunately, this approach would have the disadvantage of the relatively small insert size of such vectors, and as such would require a far larger input of resources than the remainder of the mapping effort. Thanks to this constraint, such an approach has not been tried in this current work; but as a future work would be the obvious path to take.

Chapter 5: Inversions and duplications

5.1 A duplication on the short arm of the Y chromosome

It will be recalled from chapter 4 (figure 4.1 and table 4.1) that the series of probes AA15e – 52d/C – TSPY2 located on the alphoid contig, and found in the YACs 692b1 and 872e10, also appeared to be present in the *AMELY* contig YACs 752h1 and 894f6. It was decided to investigate these probes further, since, as the YACs did not seem to overlap, a duplication was suspected.

In order to study the organisation of sequences on these YACs, complete digests were made of each of these four YACs using the same enzymes as were used to make the restriction maps (i.e., *Bsi*WI, *Bss*HII, *Mlu*I, *Sac*II, and *Sfi*I). A fifth YAC, 765g11, was also included in the digests. This YAC was from the same region as the *AMELY* YACs, although it had not been mapped comprehensively, as the region was covered in the other two YACs. The digests were then size-fractionated by pulsed-field gel electrophoresis, the DNA transferred to a nylon membrane by Southern blotting, and then hybridised with each of the probes in turn. The probes used were: i. Total human DNA, in order to visualise all fragments of the YACs. ii. The left and the right YAC vector arms, in order to eliminate any size differences between fragments solely due to the presence of the other probes under study on an end fragment of the YAC. iii. Alphoid and AMY8e probes were used, as these did not appear to be part of the suspected duplication, and so would confirm that the YACs did actually belong to the alphoid and *AMELY* contigs, respectively. iv. Finally, the filters were probed successively with the AA15e, 52d, and TSPY probes which were involved in the duplication.

The hybridisations are shown in figures 5.1 – 5.8, over the following pages.

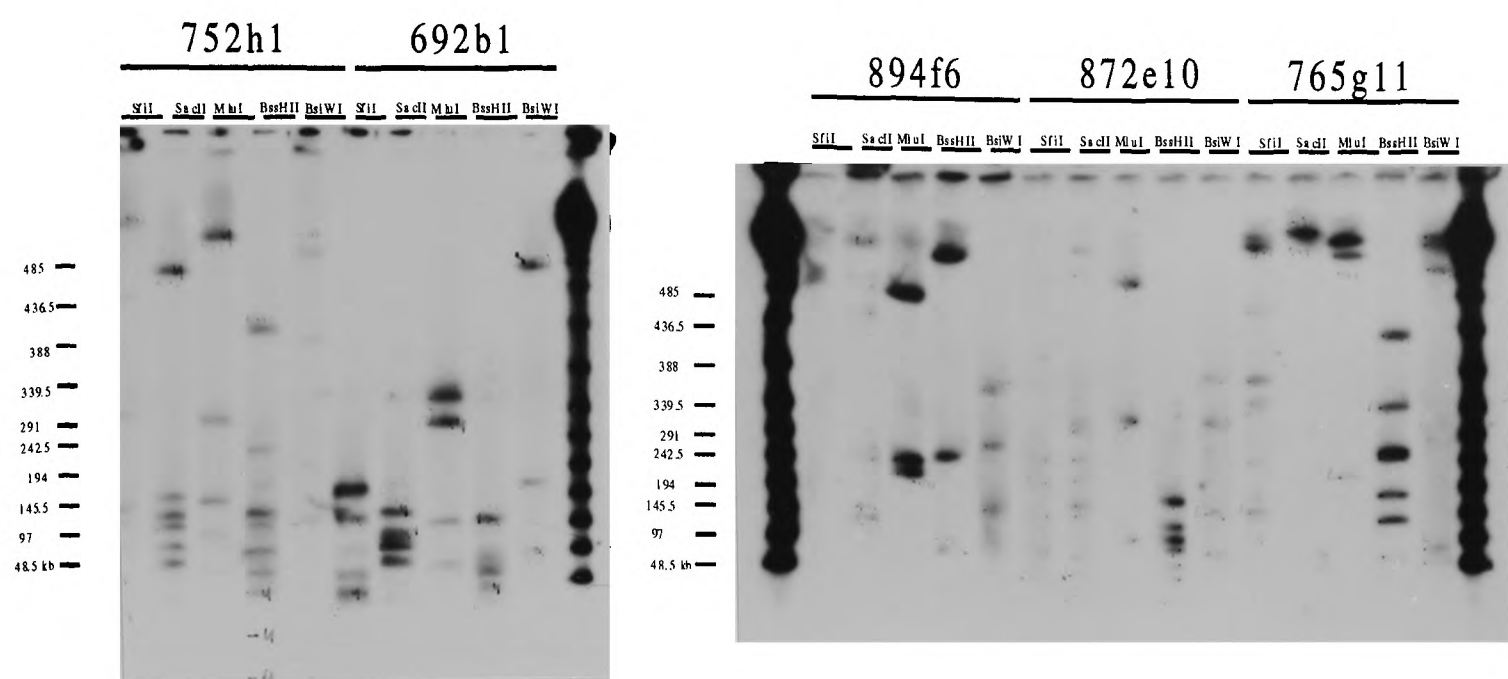


Figure 5.1: Total digests of the named YACs were hybridised to radiolabelled total human DNA, in order to visualise the fragment sizes.

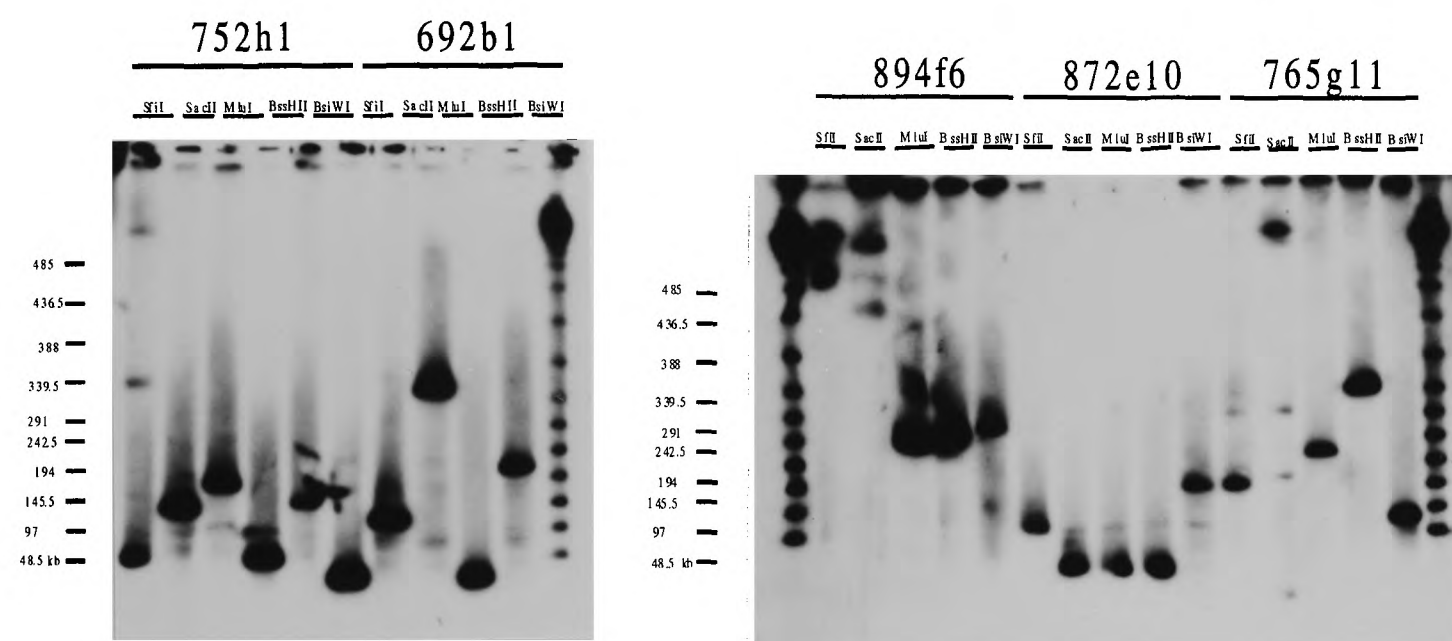


Figure 5.2: Total digests of the named YACs were probed with the left vector arm of the YAC.

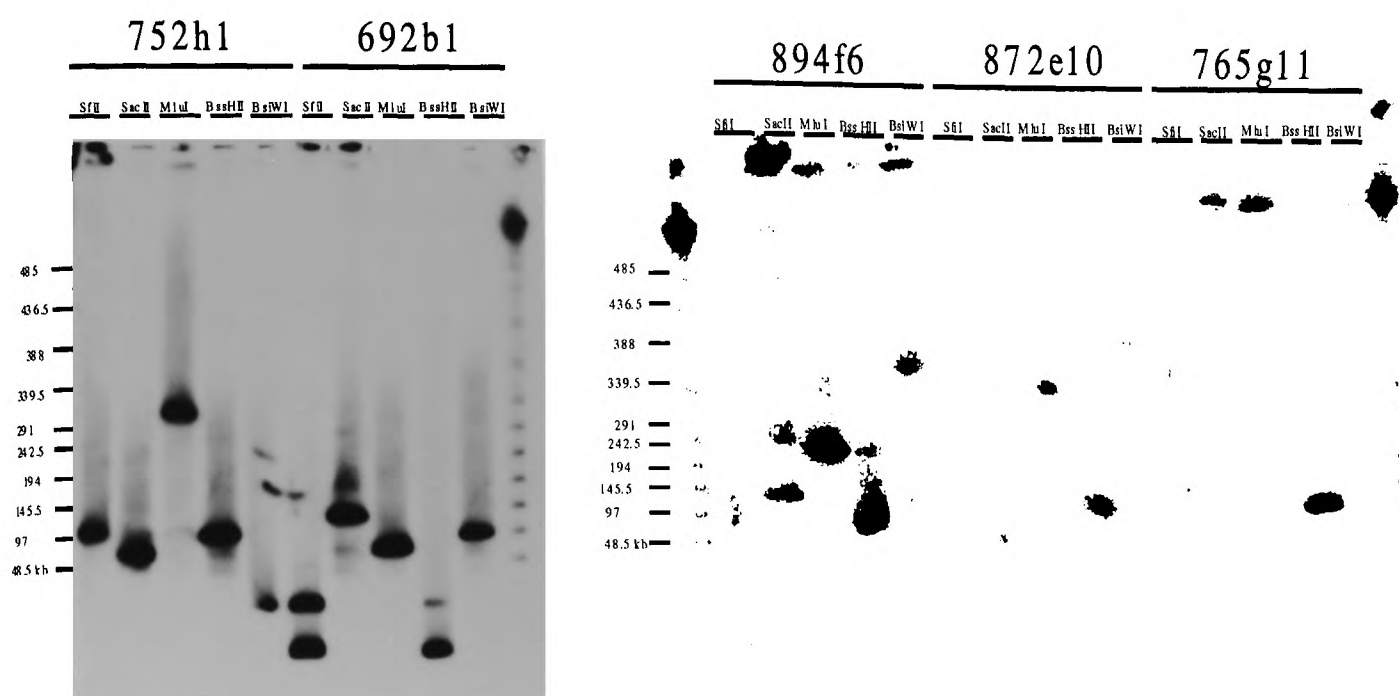


Figure 5.3: Total digests of the named YACs were probed with the right vector arm of the YAC.

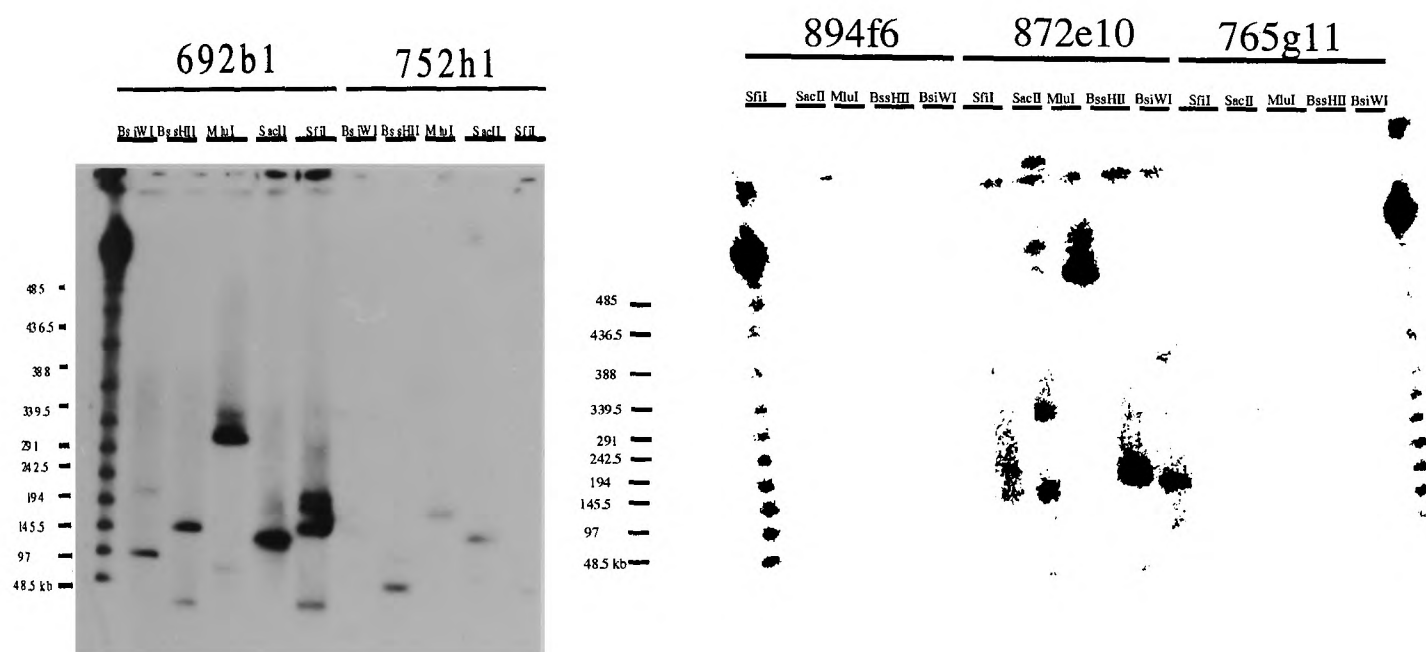


Figure 5.4: Total digests of the named YACs probed with alpha satellite DNA.

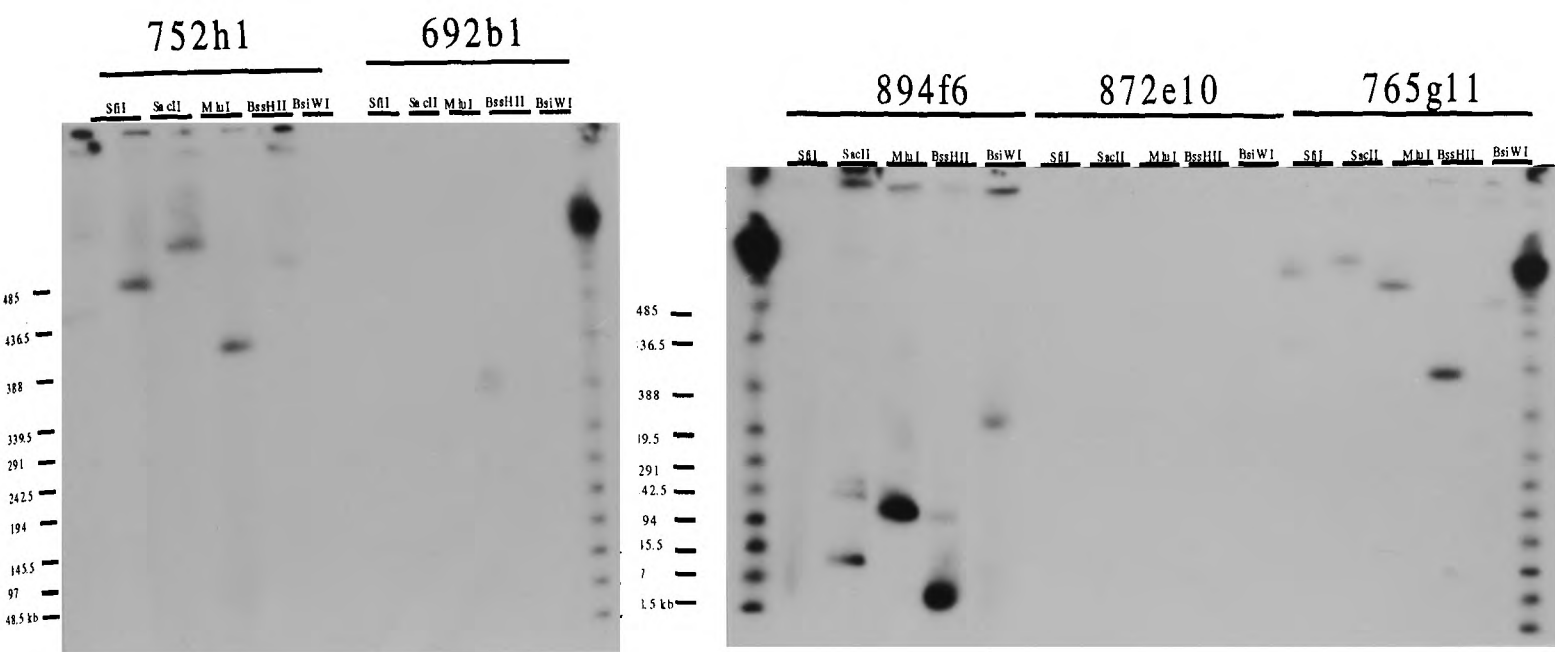


Figure 5.5: Total digests of the named YACs probed with AMY8e.

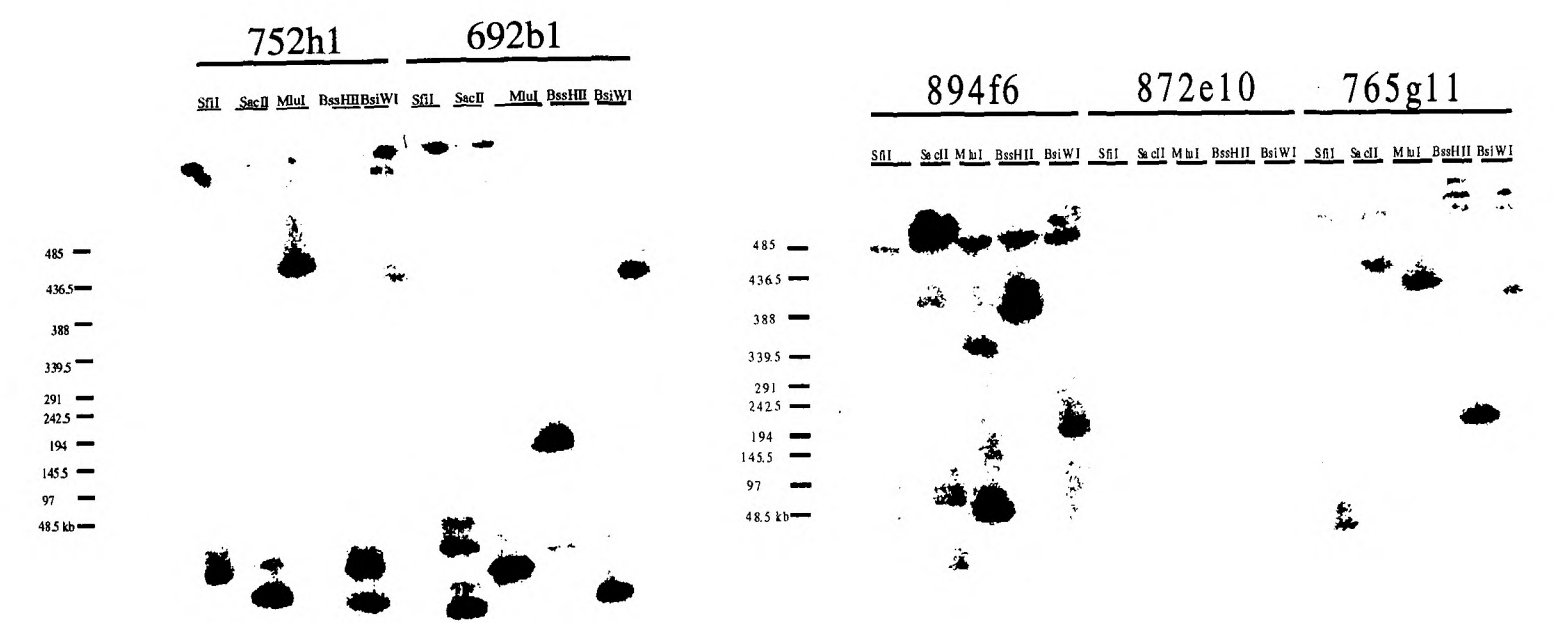


Figure 5.6: Total digests of the named YACs were probed with AA15e.

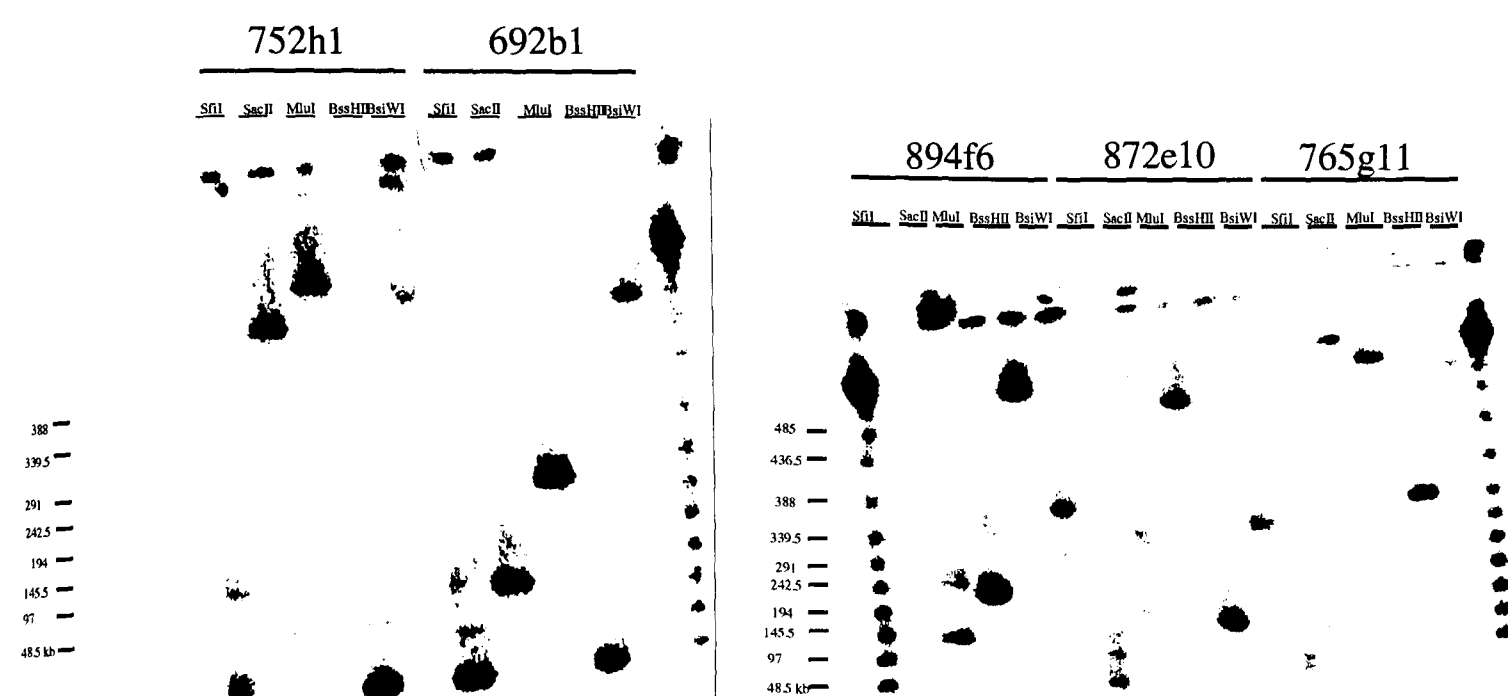


Figure 5.7: Total digests of the named YACs were probed with 52d.

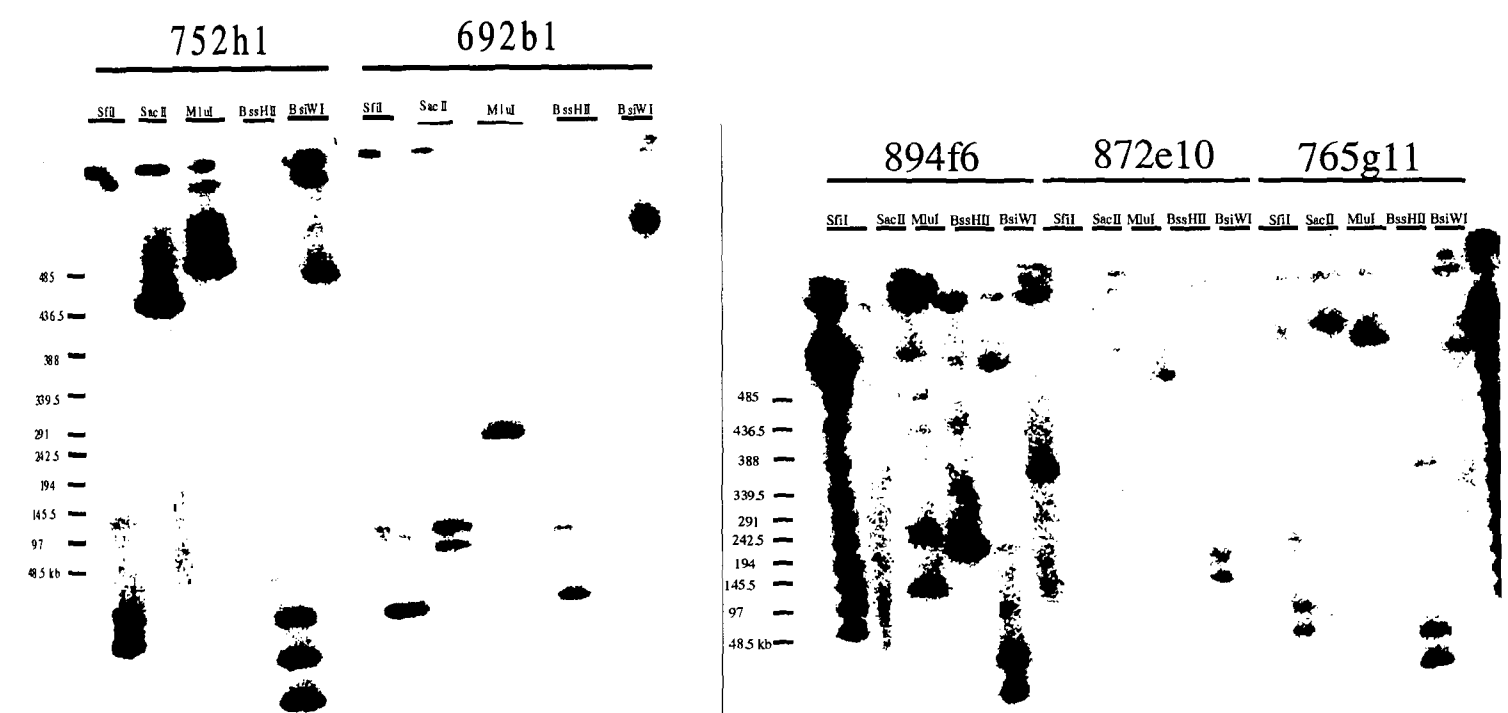


Figure 5.8: Total digests of the named YACs probed with TSPY.

The information from each of these digests was applied to the contig consensus maps, and enabled the precise localisation of each of the probes studied. An extract from the

consensus maps with each probe fragment shown, and the localisation of the probes, is shown in figure 5.9 below.

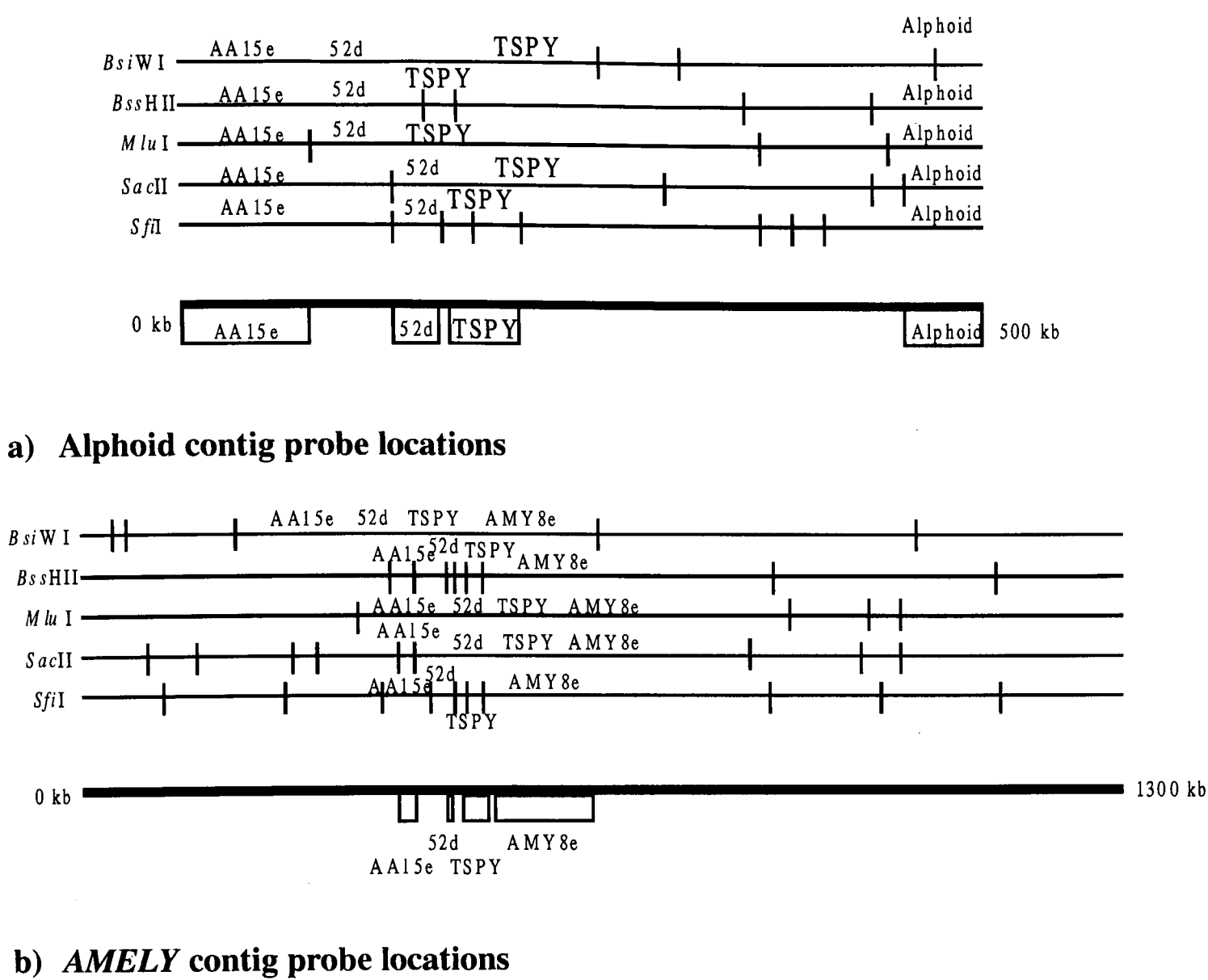


Figure 5.9: The locations of the probes as determined from figures 5.1 – 5.8. Top: The alphoid contig consensus map. Bottom: The *AMELY* contig consensus map.

The analysis shows that the block of probes AA15e – 52d – TSPY is present on both sets of YACs, although there are noticeable differences in the restriction sites around each of the probes. Note also that the complete digest allowed a number of small fragments to be seen which were not visualised when the partial digests were mapped, due to the level of resolution of the partial digests. These sites have been placed on the consensus maps with as much precision as is possible. The final locations of the *AMELY* duplication are shown on the final *AMELY* contig map, given as figure 6.10 in section 6.1; the location of the alphoid contig probes was already known, and this experiment confirmed the previous data.

As this seems to have uncovered another block of the TSPY repeat, different in restriction enzyme structure from the previous two, this block on the *AMELY* contig

will now be referred to as TSPY3. The sizes of the three blocks decrease from TSPY1 (c. 500 kb, although variable between individuals), to TSPY2 (c. 50 kb), with TSPY3 being the smallest, at around 30 kb; although unpublished data (C. Tyler-Smith, pers. comm.) suggest that the sizes of the TSPY2 and 3 blocks vary between cell lines, being 55 kb and 30 kb respectively in the OXEN cell line, and 33 kb and 50 kb in the CEPH cell line.

The discovery of this third TSPY block, as well as the duplicated nature of 52d/C and AA15e, helps to explain some of the discrepancies between the various maps which have been assembled of this region. With respect to mapping studies, the existence of a previously unrecognised duplication suggests that the contigs assembled in other studies may not be true contigs after all, but rather unite YACs from these two separate regions – recall that judging solely by probe content (or STS content), then these two sets of YACs seemed to overlap, so linking the alphoid and *AMELY* contigs. Combined with previous knowledge about the non-uniqueness of certain of the STSs used in these studies, this duplication is further evidence against the reliability of these maps.

This also affects the localisation of a number of genes from the chromosome. Several genes were reported in the second Y chromosome workshop (Affara et al., 1997) as being localised with respect to the two original blocks of TSPY (for example, *AMELY* itself, *PRKY*, and *GBY* [a genetic function, rather than a gene]). This new data does not directly contradict the placement of these genes on the chromosome, but suggests that the level of precision with which they were placed is not as high as previously thought.

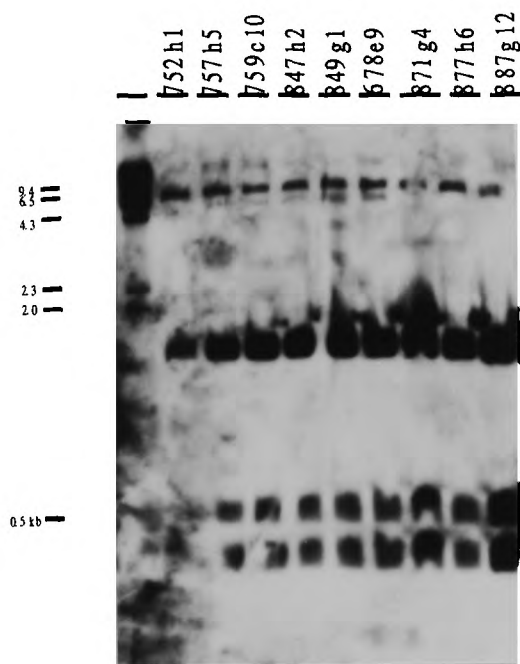
A final important aspect of this duplicated region is that it may turn out to be of importance in Y-chromosomal rearrangements on the short arm. Evidence of these rearrangements, and the possible role of the duplication in them, is presented in section 5.3.

5.2: An X-Y homologous region on the *AMELY* contig.

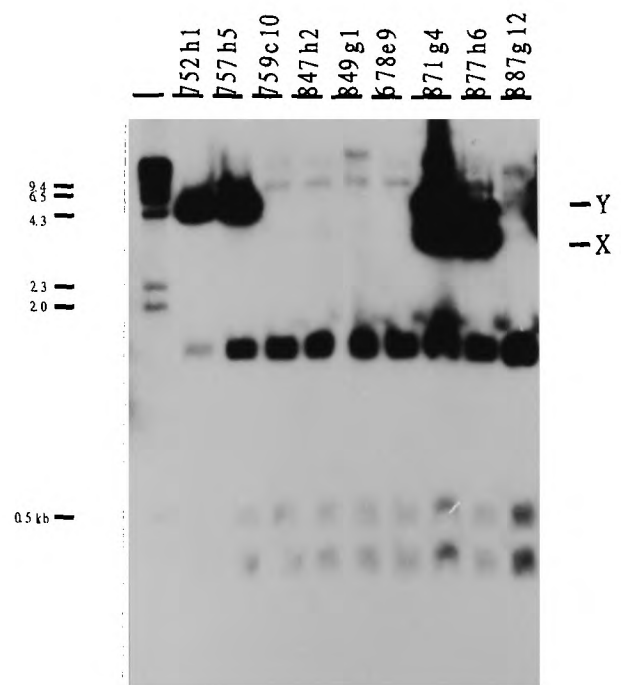
Elsewhere on the *AMELY* contig, the YACs 759c10, 804f3, 847h2, and 849g1 had all been mapped, and seen to extend the region covered by some 700 kb. As this was now the extreme end of the contig (and, if the orientation of the contig was such that this end is the most distal, also the extreme point of the whole mapping study), it was decided to investigate the nature of the sequences in this region, in order to try to relate this map to other studies of the Y chromosome. As can be seen in figure 4.10, the end-rescued probe 752e not only detects one locus on the Y, but also another locus elsewhere in the genome, either X chromosomal or autosomal.

In addition to this, there were also a number of other YACs which had been identified as containing 752e, but no other probes (YACs 678e9, 871g4, 877h6; see also table 4.2). It would therefore be of interest to see if these YACs did lie on the *AMELY* contig, or were autosomal / X chromosomal.

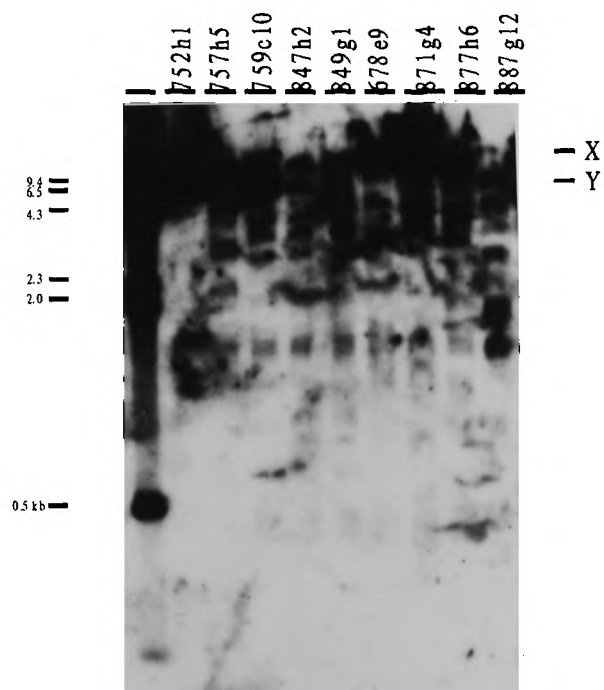
It was judged most likely that the 752e probe would be X-Y homologous, rather than Y-autosomal, since the X-Y homologous region covered by *AMELY* – pDP34 – GMGXY8 was known to lie close to a larger more distal block of X-Y homology. To this end, a set of X-Y homologous probes (GMGXY4, GMGXY5, GMGXY9, and GMGXY12) was obtained from C. Sargent, Cambridge University. All the YACs from this part of the *AMELY* contig, and the three unknown 752e YACs were digested with *TaqI*, size-fractionated on a conventional gel, transferred to a nylon membrane, and probed sequentially with each of these probes. An additional YAC tested was 887g12 (see section 4.4), as the nature of this YAC was still unknown at this time, and so it was decided to include it in the analysis. The X-Y homologous probes detect *TaqI* fragments of differing sizes, depending on whether the fragment is X- or Y-chromosomal in origin; other bands in the figures common to all lanes are the result of vector hybridisation from the plasmid probe. The hybridisations are shown in figure 5.10, below, and the results summarised in table 5.1.



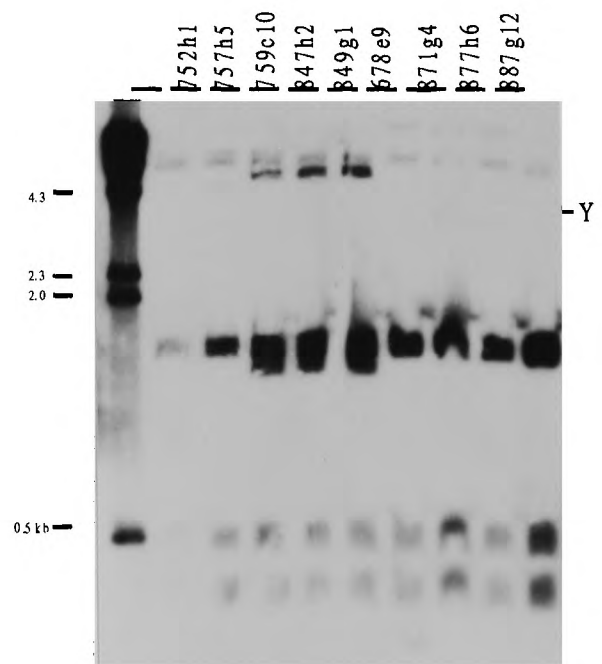
a) GMGXY4



b) GMGXY5



c) GMGXY9



d) GMGXY12

Figure 5.10: *TaqI* digests of the 752e YACs probed with GMGXY4, XY5, XY9 and XY12. The results are summarised in table 5.1.

Table 5.1: The probe content of the 752e YACs				
	GMGXY4	GMGXY5	GMGXY9	GMGXY12
752h1	-	Y	Y	-
757h5	-	Y	Y	-
759c10	-	-	Y	Y
847h2	-	-	-	Y
849g1	-	-	-	Y
678e9	-	-	X	-
871g4	-	X	X	-
877h6	-	X	X	-
887g12	-	-	-	-

It can be seen from the figure and the table that those YACs which have been mapped to the *AMELY* contig all contain the Y-chromosomal loci of these probes; while the unmapped YACs are all of X-chromosomal origin. Confirming the results of section 4.4, 887g12 is not from this region, on the X or the Y chromosome.

The results also allow us to place GMGXY5, 9, and 12 on the contig map (see the final map in section 6.1, figure 6.10). According to the main study of these X-Y homologous probes (Sargent et al., 1996), all these probes are from the large X-Y homologous block of Yp / Xp21.3. GMGXY4, not found in any of the YACs mapped in the current study, is placed distal to the other three probes, and therefore is expected to lie outwith the region under investigation. The order of the other probes along our contig, and the distal location of GMGXY4, confirms the orientation of the *AMELY* contig, with this end extending toward the telomere, and the 758e side of the contig being closest to the TSPY contig.

Linking the current map with previously published studies, this large X-Y homologous region has been estimated as covering some 2.8 Mb (Sargent et al., 1996), extending almost as far as the *ZFY* gene on Yp, which lies immediately adjacent to *SRY* and the PAR1. This information anchors our map firmly with respect to other Y-chromosome studies.

5.3: An inversion polymorphism on the Y chromosome short arm.

Thanks to a number of discrepancies between the different maps, and particularly some anomalous results in deletion panel mapping of the Y chromosome, the existence of an inversion polymorphism on the short arm of the chromosome had been long suspected. A number of groups have independently found evidence for an inversion polymorphism (Vergnaud et al., 1986; Affara et al., 1986b, Tyler-Smith et al., 1988; Sargent et al., 1996). For example, Tyler-Smith et al. (1993) analysed the probe content of a number of Y chromosomes with terminal deletions. One subclass of deleted chromosomes appeared to have, in addition to a terminal deletion, an internal deletion. One explanation for this internal deletion may be that an inversion had transposed this interval so that it lay more distal in this subclass, and so was included in the terminal deletion. The other subclass of chromosomes did not lose this interval as it was not inverted, and so lay internally to the terminal deleted region.

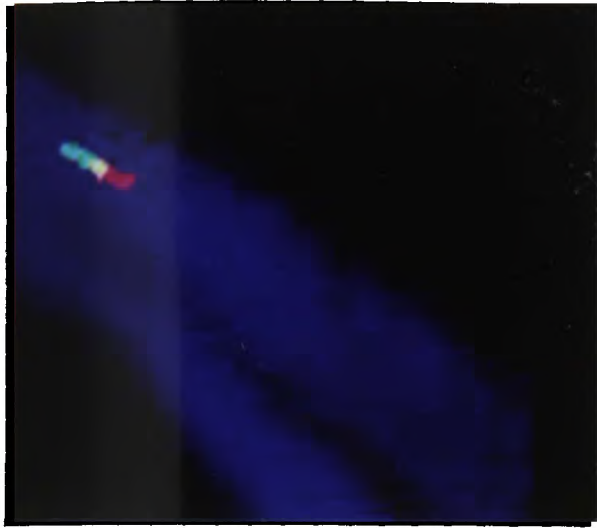
Evidence from the physical YAC mapping comes from the ordering of the contigs – it will be recalled from section 4.4 that the YAC 887g12 linked our TSPY and alphoid contigs; however, the original Foote et al. map suggested that the order of STSs in this region would lead to the contigs lying in the order alphoid – *AMELY* – TSPY. The ‘missing’ interval would appear to cover the smaller X-Y homologous region on Yp, and include the probes GMGXY8 and *AMELY*. The deletion intervals which flanked this region corresponded to the TSPY and alphoid contigs.

It was decided to try to investigate this proposed inversion by means of FISH mapping – this would give an unambiguous visualisation of the order of the probes on genomic DNA. Standard FISH techniques on metaphase chromosomes were discounted as a possible technique, as these techniques do not give a high enough resolution to order the regions on the chromosome; although metaphase FISH was used in order to check that the probes used hybridised to the expected region of the Y chromosome. Thus, methods for producing high-resolution FISH were investigated.

After examining several different techniques (see Heng, Squire & Tsui, 1992; Haaf & Ward, 1994a and 1994b; Parra & Windle, 1993; Heiskanen et al., 1994), the method detailed in Haaf & Ward was chosen as yielding images of the appropriate resolution. This technique relies upon mechanically stretching chromosomes by means of centrifugation onto the microscope slide, rather than chemically releasing the DNA, as used by other methods.

Two commercially available probes were used to investigate the inversion, both obtained from Oncor. DYZ3 detects the alphoid repeat, and was labelled with biotin, and detected with FITC-avidin, giving a green signal. The *AMELY* probe was labelled with digoxigenin, and detected with rhodamine-anti-digoxigenin, giving a red signal. The third probe used was a plasmid containing part of the TSPY1 repeat; this was labelled with both biotin and digoxigenin, thereby giving a third colour, yellow, when combined with the FITC and rhodamine detection steps. The chromosomes were counterstained with DAPI, giving a blue colour.

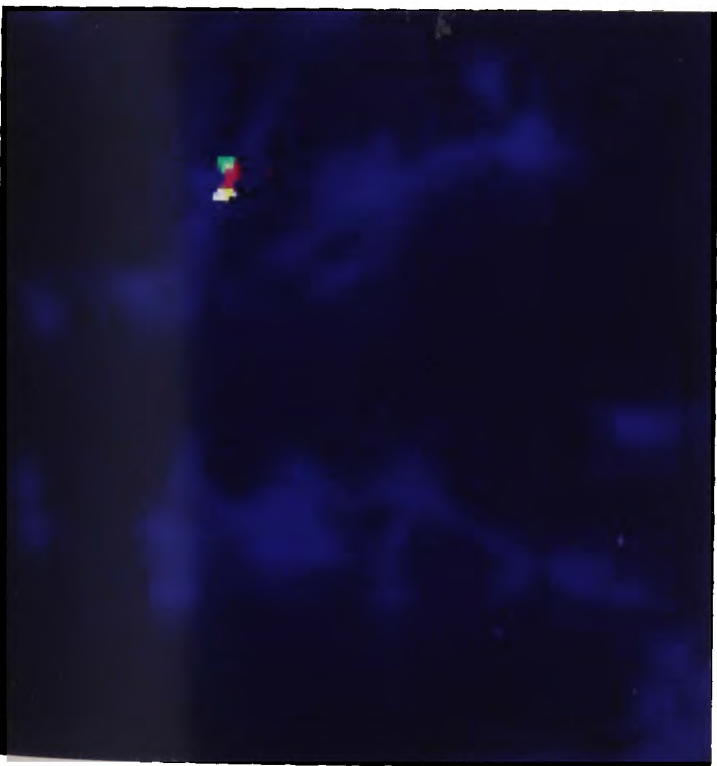
Three lymphoblastoid cell lines were investigated: Boleth, a 46, XY male, and the source of the CEPH YAC library; OXEN, the 49, XYYYYY male source of the OXEN cell line; and M100, a normal 46, XY male cell line with the same Y haplotype as OXEN (Pandya, 1998). Sample images from each of the three cell lines are shown below, in figure 5.11.



a) The Boleth cell line



b) The OXEN cell line



c) The M100 cell line

Figure 5.11: FISH images of three cell lines probed with DYZ3 (green), *AMELY* (red), and TSPY1 (yellow). The chromosomes were counterstained with DAPI (blue).

It can be seen from these images that the order of the probes differs between the cell lines. The Boleth cell line has the order alphoid – TSPY – *AMELY*; while the OXEN and M100 cell lines are of the form alphoid – *AMELY* – TSPY. This shows that an inversion polymorphism does exist between different Y chromosomes.

The number of cells studied was low. The stretching technique gave a relatively low number of interpretable signals per slide. The degree of stretching of the chromosomes was uneven, even within a single slide, so a large number of chromosome spreads were either too extended or not extended enough to obtain an image. This effect can be seen in figure 5.11b, the image of the OXEN cell line. As this is a 49, XYYYYY cell line, it would be expected that there would be four visible hybridisations, one for each Y chromosome. However, only two can be seen in this image, suggesting that the other two Y chromosomes were not extended suitably for hybridisation to occur.

Nonetheless, several interpretable images were obtained from each cell line. In total, 11 images were analysed from the Boleth cell line, 7 from the OXEN cell line, and 6 of the M100 cell line. All these images were also obtained from at least two separate cell cultures and slide preparations (three in the case of the Boleth cell line). All of the images were consistent with the orientation determined from the figure above. Interpretable images were defined as those which showed three distinct signals on the Y chromosome; uninterpretable images either showed less than three signals, or the signals overlapped, or the order of the probes could not be determined (i.e., if the chromosome was too extended then a single line could not be followed along the signals). This confirms that the inversion is indeed a true polymorphism.

What is the frequency of the occurrence of this inversion? Obviously, the current sample size of three is not large enough to give any answers. However, it is suspected that the form represented by OXEN (alphoid – *AMELY* – TSPY) is the more common, at least among deleted chromosomes. The evidence from this comes from the deletion mapping studies referred to earlier. In the study by Sargent et al. (1996), 57 deleted Y chromosomes were studied; of these, 22 chromosomes had deletions in the informative region, from the alphoid to the TSPY1 intervals. Of these 22, 7 individuals had what seemed to be an internal deletion of the *AMELY* interval – i.e.,

these seven individuals had the alphoid – TSPY – *AMELY* ordering. None of the individuals studied in the report by Vollrath et al. (1992) showed evidence of having this ordering. Of course, further studies directed toward answering this specific question would yield more definite answers, particularly as the individual chromosomes studied in these reports were noteworthy due to the fact that they had deleted part of the chromosome, and so may be prone to having other rearrangements of the chromosome. Alternatively, it may be that possession of the probe order alphoid – TSPY – *AMELY* somehow predisposes chromosomes against undergoing further rearrangements, and so studies of deleted chromosomes would significantly underestimate the prevalence of such a probe order.

The significance of this inversion lies mainly in the field of mapping the chromosome. It will be recalled that, while the Foote et al. map was constructed from an OXEN-based YAC library, the Jones et al. map, on the other hand, was based on the CEPH YAC library. The current map is based on YACs from both libraries, and so would be expected to mark the inversion breakpoints somewhere on the map. However, this has not been found to be the case – both sets of YAC maps agree as far as the mapping has been carried out. In all likelihood, this merely reflects the fact that the coverage of the map in the OXEN library is not as extensive as in the CEPH library, and the actual breakpoints have not been intensively mapped.

An important question to answer is where do the inversion breakpoints lie? As mentioned above, no differences in the map between different YAC libraries have yet come to light. One likely place for the breakpoint to occur is in the duplicated region described in section 5.1, with copies on the alphoid and the *AMELY* contigs. Given the orientation of the contigs determined thus far, these two duplicated blocks appear to form an inverted repeat, flanking the region which contains *AMELY* and TSPY1 (which are the two sequences known to be involved in the inversion from the FISH evidence). That is to say, the order of the sequences along the chromosome is alphoid – TSPY2 – 52d/C – AA15e – TSPY1 – *AMELY* – AA15e – 52d/C – TSPY3. (Of course, the order of *AMELY* and TSPY1 will be reversed in one group of Y chromosomes.) This inverted repeat may well mediate the inversion of the entire intervening block of sequence; this possibility is considered further in chapter 7. Such a mechanism leading to polymorphism is uncommon in humans, as autosomal

inversions will create problems with recombination between homologous chromosomes, leading to duplications or deletions rather than a stable inversion polymorphism. However, at least one other such region is suspected to be involved in an inversion polymorphism. The emerin gene associated with Emery-Dreifuss muscular dystrophy on the X chromosome is flanked by inverted repeats, and the region is found to be inverted in around 20% of X chromosomes (Small & Warren, 1998). In this case, the region involved in the inversion is only around 48 kb, compared with the postulated inverted region on Yp covering some 4 Mb. Nonetheless, it seems likely that such a mechanism may be found acting on the Y chromosome to create this inversion polymorphism.

The emerin inversion can be detected by a restriction fragment length polymorphism, giving different restriction fragment sizes for a certain region depending on the orientation of the emerin gene. However, such a simple diagnostic tool has not yet been discovered for the Y chromosomal inversion. Development of the test would certainly be of benefit, in that it would make determination of the orientation of the inverted region easier, and would also help to localise the inversion breakpoints more precisely.

A further consequence of the existence of the inversion is in the field of gene mapping. A recent report (Gläser et al., 1997) used similar FISH techniques to the ones in this section to order a number of genes along the Y chromosome. The reported order for some of these genes was TSPY1 – *PRKY* – *AMELY* – TSPY2 – alphoid. This suggests that the studies were carried out on the more common (OXEN) form of the chromosome; and that the CEPH form of chromosome would provide a different order. While not of great significance, this is important to consider when mapping data is presented.

The evolutionary significance of this inversion is unknown., but may illustrate the rapid cytogenetic evolution of the Y chromosome. The question of effects of the inversion on fitness will be addressed in section 6.3. However, the detection of a novel polymorphism is important for population studies. Several studies are ongoing, and make use of polymorphisms on the Y chromosome to determine relationships between populations. It would be instructive to analyse many samples for this

inversion to supplement the groupings already made of Y chromosomes. If the inversion arose only once, it should fit other polymorphisms in a simple way and be placed in a single location on an evolutionary tree of the Y chromosome. Alternatively, it may be the case that this inversion has occurred several times, but is more likely to occur on only a limited subset of all the Y chromosomal haplotypes, in which case further study would reveal this fact. This idea is of some relevance to the ideas and results presented in chapter 6.

Chapter 6: Genes and Mutants

6.1 Localising genes on the chromosomal map

Over the course of this work, a number of groups working on the Y chromosome reported novel genes from the area of the Y covered in this present work. In several cases, these genes had not been localised with much precision on the chromosome, beyond assignation to a broad deletion interval. It was decided that it would be of great utility for understanding the biology of this region to place these novel genes on the current map.

The first gene investigated was the *PRKY* (protein kinase Y) gene (Klink et al., 1995). This is an X-Y homologous gene, known to locate to the short arm of the Y chromosome, while the X homologue (*PRKX*) is located on the short arm of the X chromosome. The Y copy of the gene had also been mapped by means of high-resolution FISH, and placed somewhere between the *AMELY* and *TSPY1* loci (Gläser et al., 1997). This gene is also of interest as a known site of chromosome instability, being a hotspot for illegitimate homologous recombination between the X and Y chromosomes – more will be discussed on this point in section 6.3.

The second gene investigated was the *DFFRY* (Drosophila fat-facets related Y) gene. Brown et al. (1998) report the *DFFRY* gene as lying on proximal Yq. The actual function of the gene is as yet unknown, although it is of interest as it lies in the critical region for three abnormalities – Turner syndrome; GBY (gonadoblastoma); and AZFa (Azoospermia factor, locus a – see Vogt et al., 1996). The critical genes for all of these have been mapped to this region using deleted Y chromosomes. Thus, further study of this gene, and refinement of its location may well help to determine if it has any involvement in these disorders.

Finally, a recent work by Lahn & Page (1997) identified many novel Y chromosomal genes – 12 previously unknown cDNA sequences (one of which was a rediscovery of *DFFRY*) were isolated from testis tissue, and subsequently located to the non-recombining portion of the Y chromosome. Many of these putative Y chromosomal genes were localised on a deletion interval map of the chromosome; those which

appeared to lie in our regions of interest were selected for study. These genes are, on the short arm, *TTY1* and *TTY2* (testis transcript Y), with no obvious open reading frames, homologues or function, and *PRY* (*PTP-BL* related Y), a testis-specific gene with homologies to a protein tyrosine phosphatase. Several more genes were suggested to lie on the long arm, such as *DBY* (DEAD box Y), *UTY* (ubiquitous TPR motif Y), the mouse homologue of which is involved in H-Y antigen production, *TB4Y* (thymosin β 4, Y copy), the X homologue of which is involved in actin sequestration, and *BPY1* (basic protein Y), which has no obvious function or homologues.

The simplest method of localising all of these genes was by polymerase chain reaction, PCR. Primer sequences and conditions for *PRKY* were obtained from Gudrun Rappold, Heidelberg; for the remaining genes, the cDNA sequences were obtained from the GenBank database and primer sequences designed from these. The primers used in all these PCR reactions are listed in chapter 2, section 2.9. The PCR reactions were tested on genomic 46, XX and 46, XY DNA samples, under a range of conditions, to ensure that amplification of the desired fragment took place. Once the reactions were optimised, the experiments were carried out on a panel of YAC DNAs from either the short arm or the long arm. The results for all these genes are shown in the figures below.

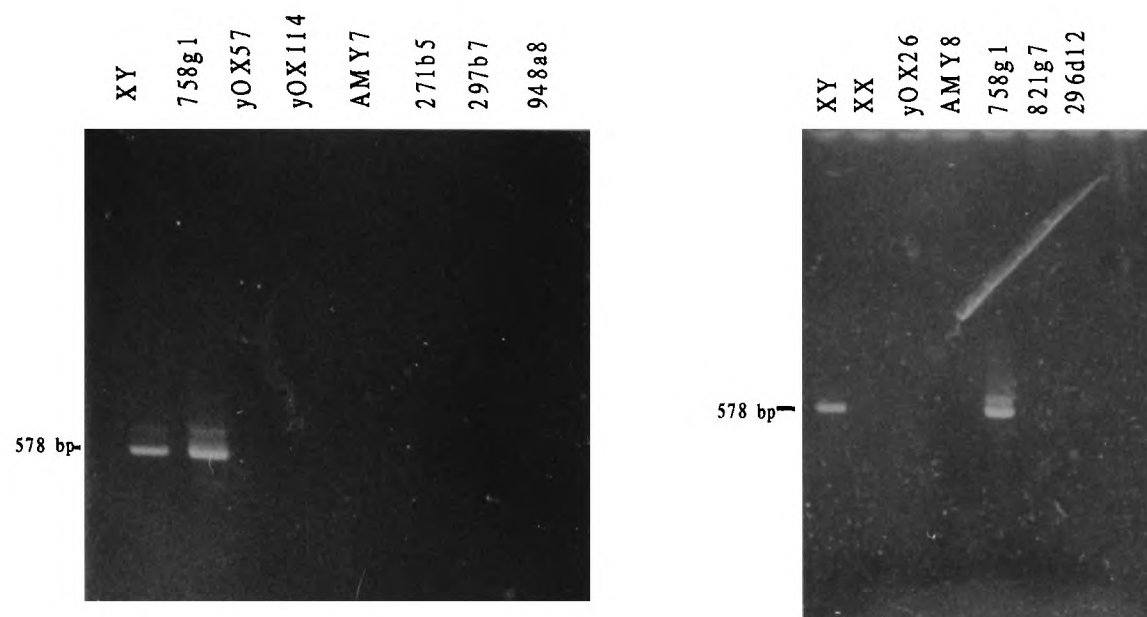


Figure 6.1: PCR of the *PRKY* gene on a selection of YACs from the short arm.

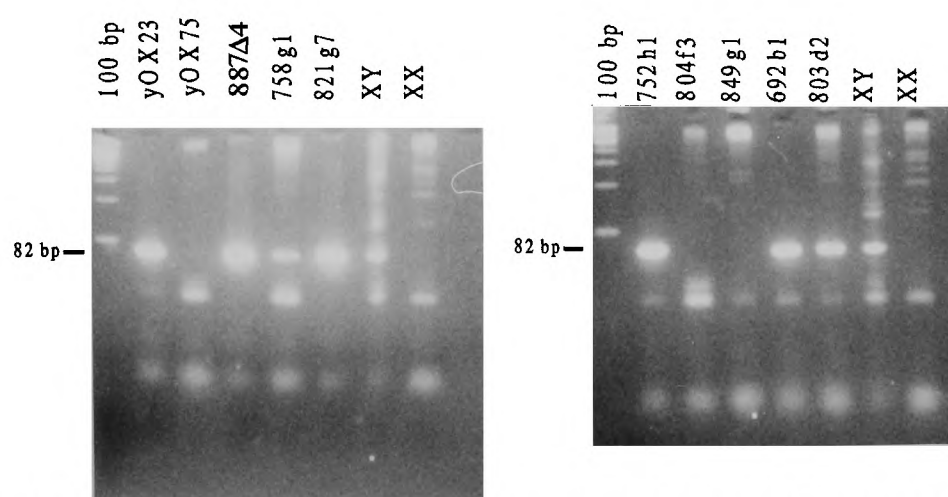


Figure 6.2: PCR of the *TTY1* gene on a selection of YACs from the short arm.

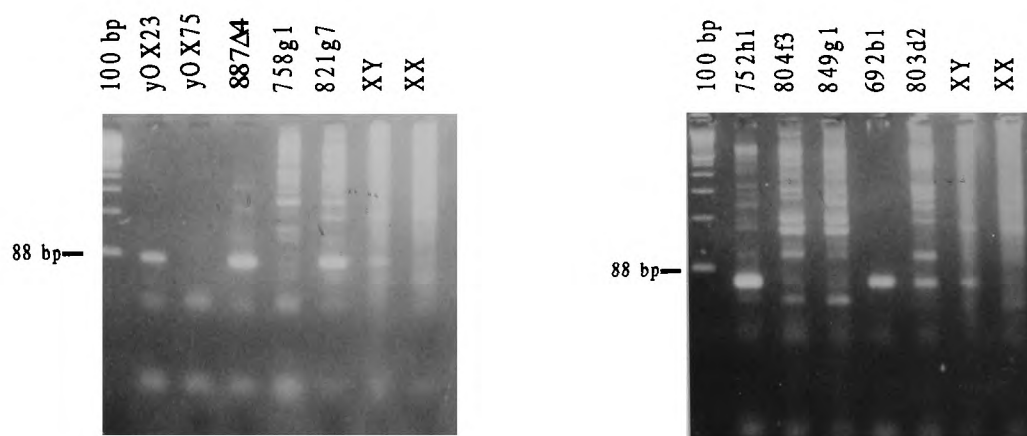


Figure 6.3: PCR of the *TTY2* gene on a selection of YACs from the short arm.

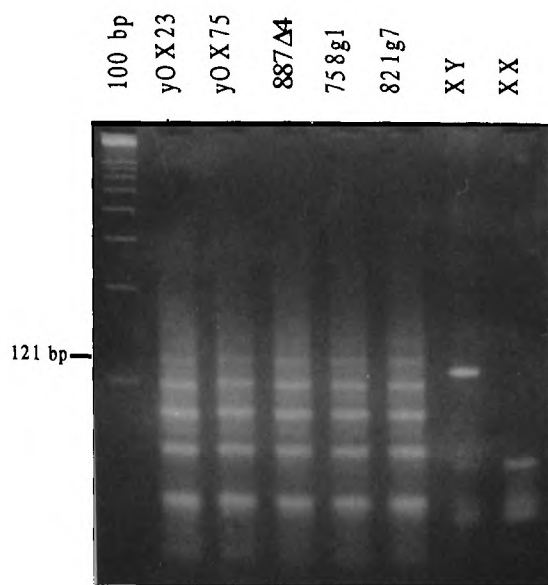


Figure 6.4: PCR of the *PRY* gene on a selection of YACs from the short arm.

This first set of PCRs, performed on the short arm YACs, localises several of the genes. *PRKY* is seen to only be present in the YAC 758g1, and none of the overlapping YACs. This shows that the gene must be placed on that part of 758g1 which is not covered by any other YACs; lying just proximal to the KFC3 probe from this region. (That is, proximal assuming the CEPH-type orientation of the chromosome.)

TTY1 and *TTY2* can be seen to localise both together, and also on only those YACs which contain any of the three TSPY blocks. However, one important difference between the TSPY YACs and the *TTY* YACs is that yOX75, coming from one side of the TSPY1 array, does not contain the *TTY* sequences. Initially, it was thought that the *TTY* sequences may have just been a case of ‘the rediscovery of TSPY’, thanks to their extensive co-localisation. A further PCR eliminated this possibility; showing that a cosmid (cY91; Tyler-Smith et al., 1988) containing a complete TSPY repeat unit did not amplify any *TTY* sequences (results not shown). So, while the *TTY* genes can be shown to be closely associated with the TSPY repeats, they are not equivalent to TSPY; also the lack of amplification from yOX75 suggests that *TTY* sequences must flank one side of the TSPY1 array.

The final gene investigated from the short arm was the *PRY* gene. This was not found on any of the YACs tested from the short arm. Lahn & Page placed the gene in the same interval as the *TTY* and TSPY genes, and so it would have been expected to be

present in this sample of YACs. However, according to their work, *PRY* could also be found in two locations on the long arm (outwith the coverage of the present map). It may be the case that the three possible *PRY* locations are actually only related sequences, and the primers chosen only detected the long arm copies, the short arm lacking the stretch of DNA amplified by these primers. The alternative explanation is that the short arm location falls in the remaining gap on the short arm map, between the TSPY and *AMELY* contigs.

The long arm genes were investigated next, and the PCRs are shown over the following pages.

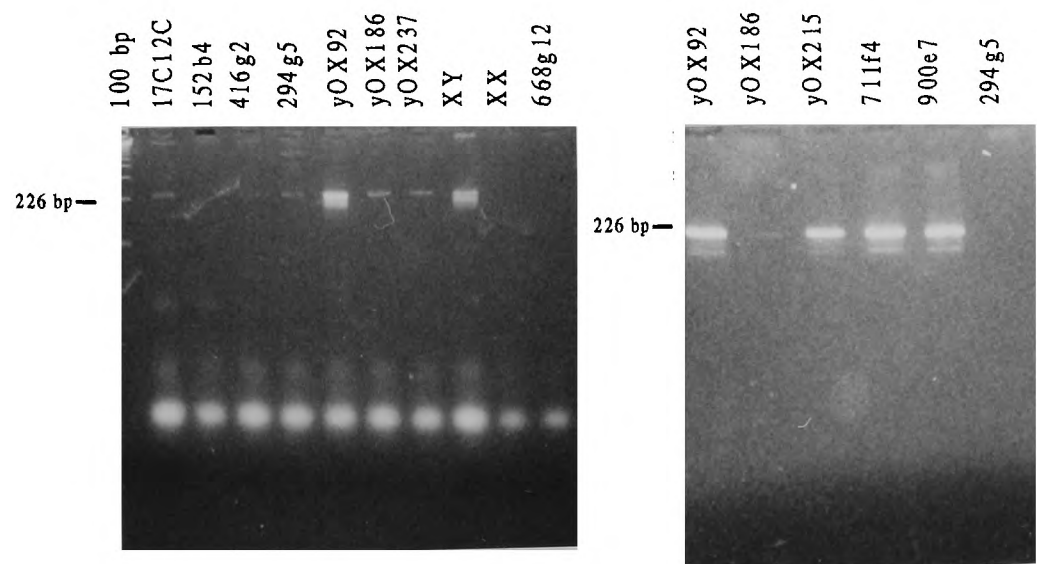


Figure 6.5: PCR of the *DFFRY* gene on a panel of YACs from the long arm.

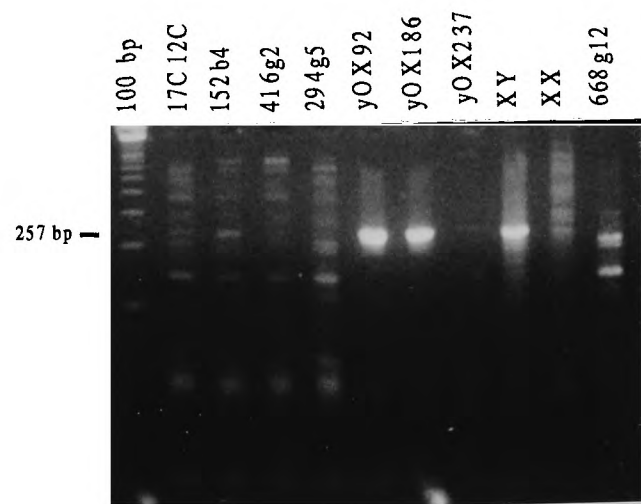


Figure 6.6: PCR of the *DBY* gene on a panel of YACs from the long arm.

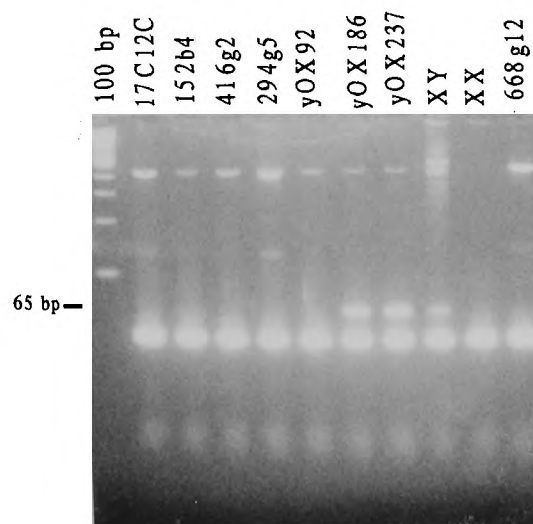


Figure 6.7: PCR of the *UTY* gene on a panel of YACs from the long arm.

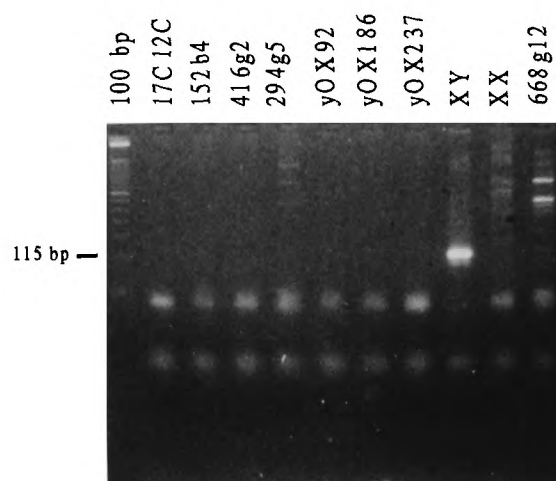


Figure 6.8: PCR of the *TB4Y* gene on a panel of YACs from the long arm.

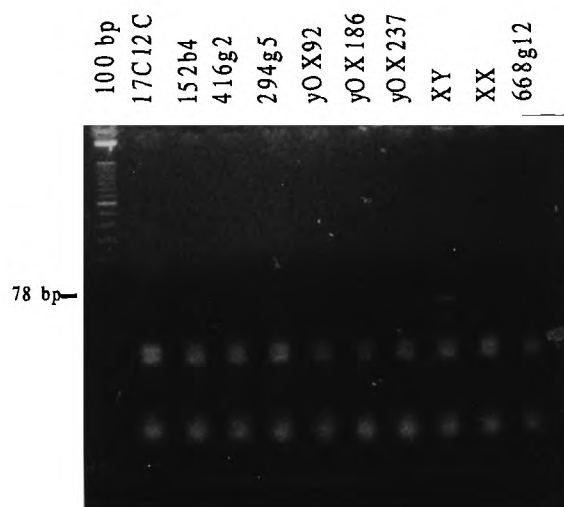


Figure 6.9: PCR of the *BPY1* gene on a panel of YACs from the long arm.

The most proximal of these five genes is *DFFRY*, which localises in practically the same place as the RBF5 probe. Which side of RBF5 the gene lies on is, however, ambiguous. The remaining genes can be placed in the order (from proximal to distal) *DBY*, *UTY*, with *TB4Y* and *BPY1* not being present on the part of the long arm

covered by this map. This order is the same as that determined by Lahn & Page (1997).

The final contig maps, with the locations of these genes marked, and the duplication described in section 5.1, are given in figures 6.10 – 6.12 over the following pages. With this additional information, we can then investigate functional and organisational aspects of the genes, in section 6.2.

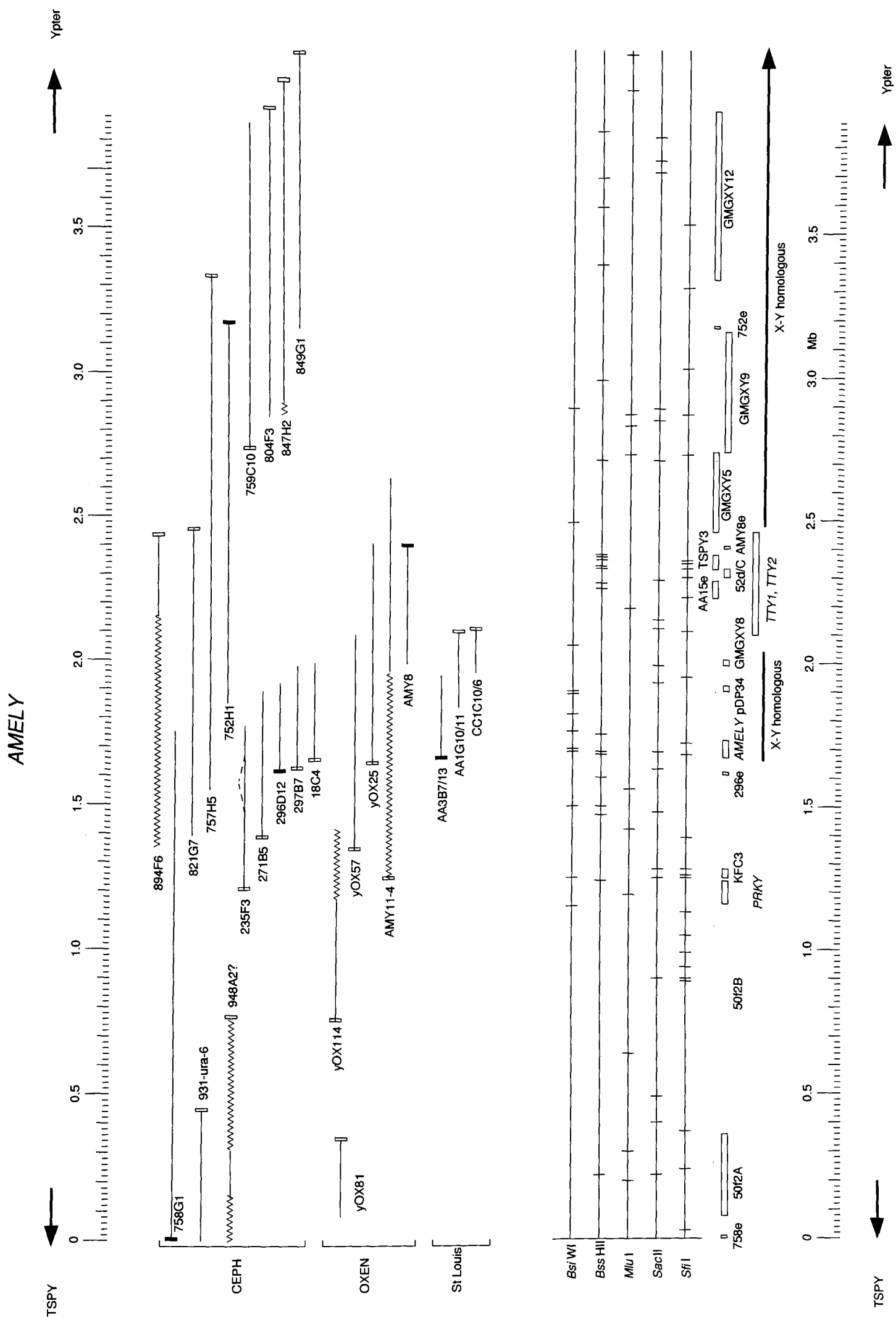


Figure 6.10: Final map of the *AMELY* contig.

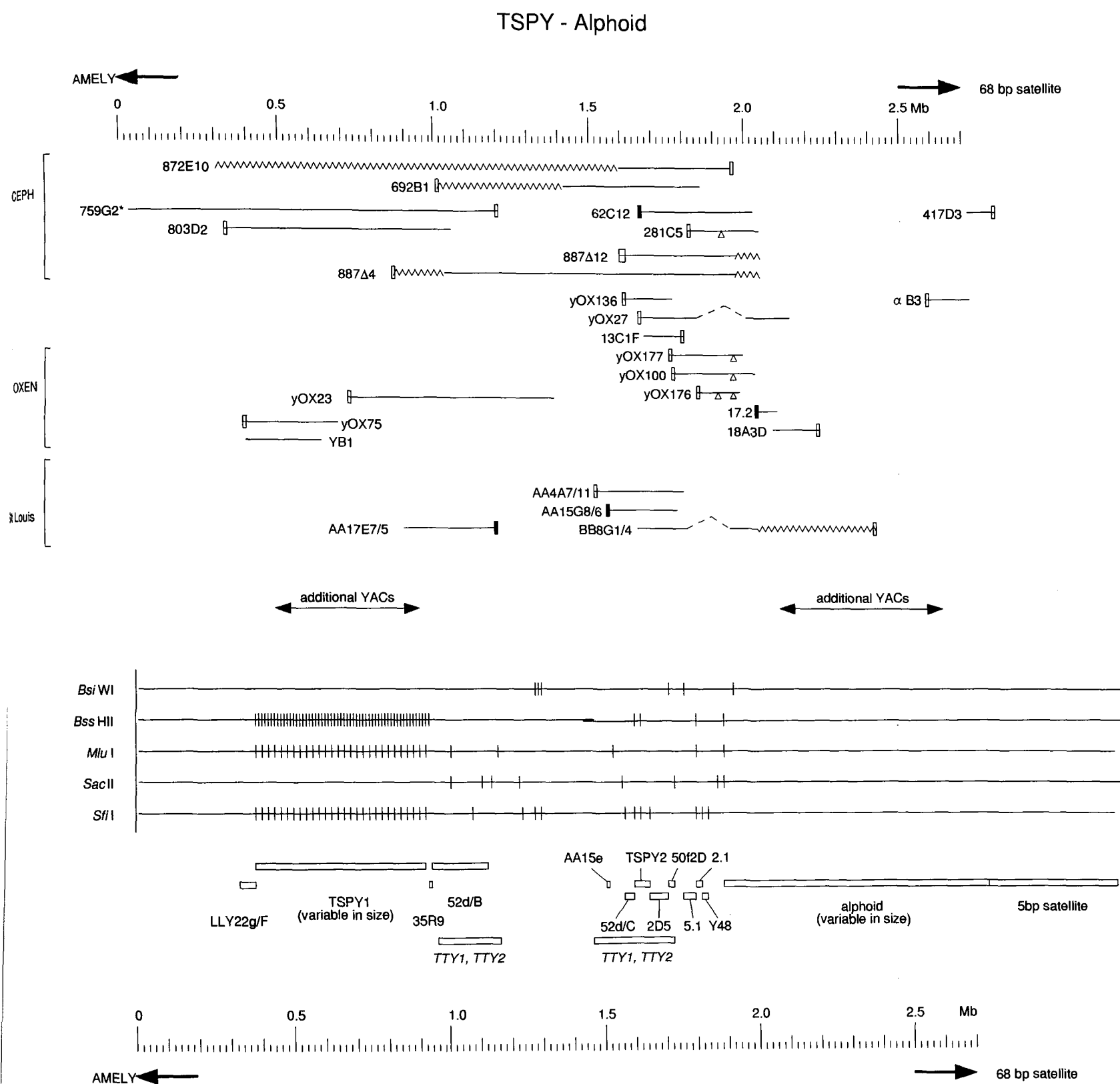


Figure 6.11: Final map of the TSPY-alphoid contig.

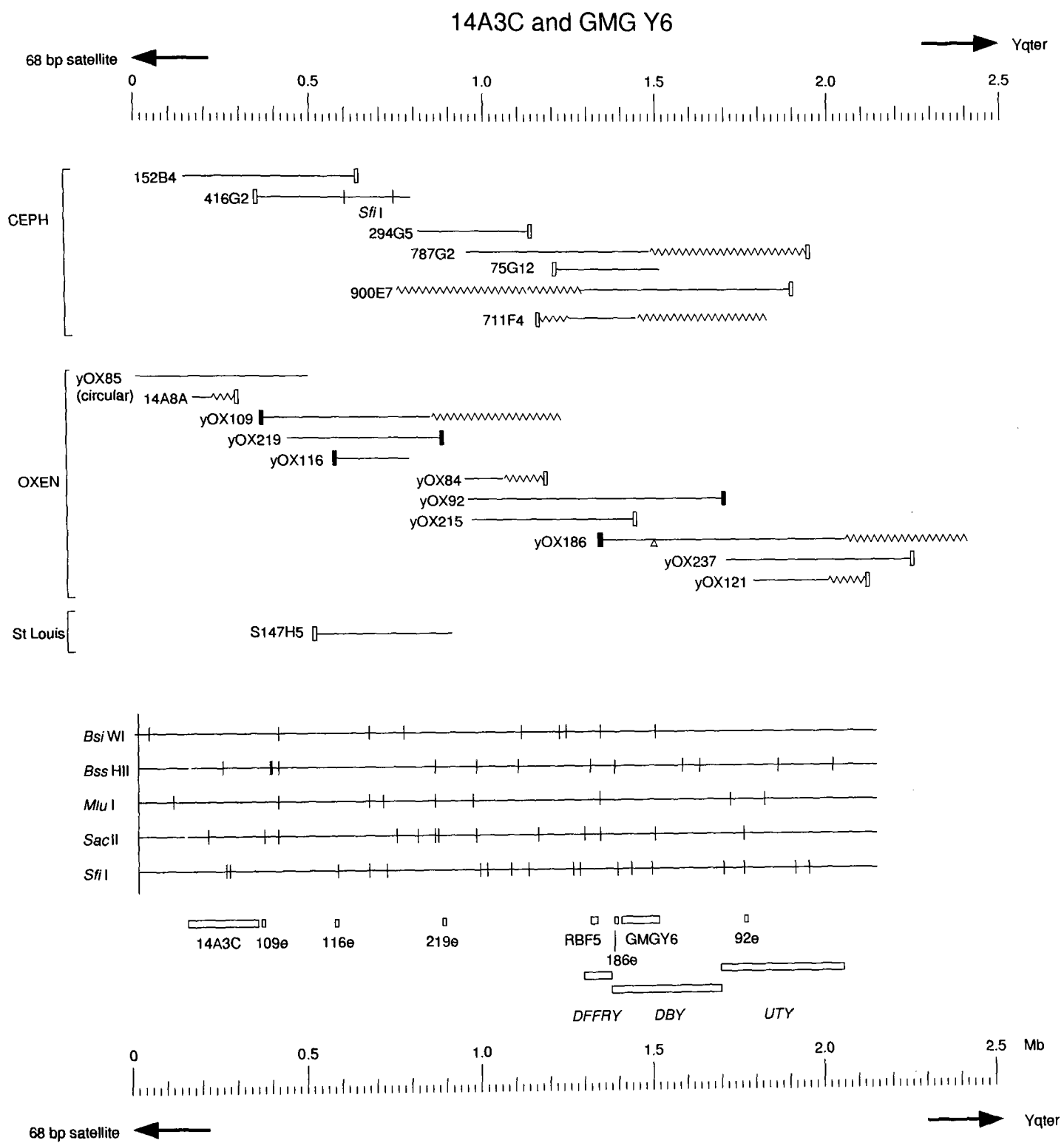


Figure 6.12: Final map of the 14A3C-GMGY6 contig.

6.2: Deleted chromosomes

Two different pairs of deleted chromosomes were investigated in this section, in order to gain additional information about the genes on the Y chromosome. Firstly, in a large-scale study (Pandya, 1998), many normal male DNA samples were tested for a large number of polymorphic Y chromosomal loci. Two apparently normal individuals from Sri Lanka were found to lack the *AMELY* locus (Santos, Pandya & Tyler-Smith, 1998). The mechanism behind this loss was suggested as being a homologous recombination between TSPY1 and TSPY2, so excising all the intervening sequences. In order to confirm this hypothesis, the individuals concerned were tested for the presence of the short arm genes localised above (i.e., *TTY1*, *TTY2*, *PRKY* and *PRY*). The results of the PCRs are shown in figure 6.13, below.

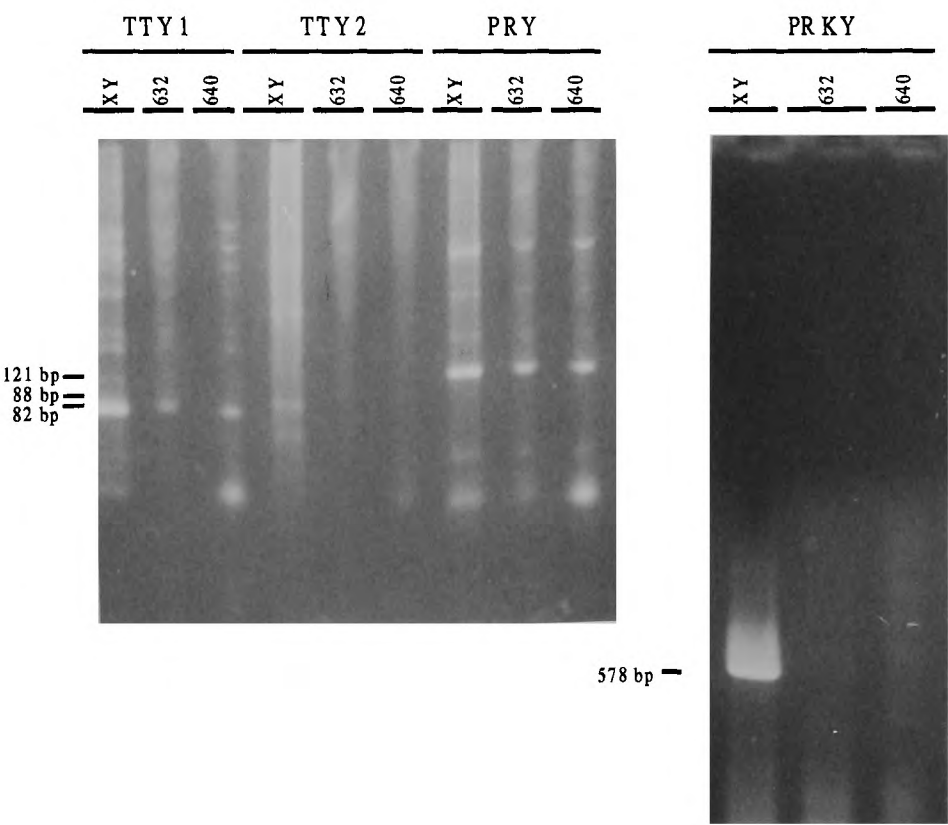


Figure 6.13: PCR of the deleted individuals (632 and 640) with *TTY1*, *TTY2*, *PRY*, and *PRKY*.

This result shows that the Sri Lankan individuals had both *TTY1* and *PRY*, while *PRKY* and *TTY2* had been deleted. These chromosomes were thought to have the

OXEN-type orientation, so giving an order of the loci of alphoid – TSPY2 – *TTY1* + 2 – *AMELY* – *PRKY* – TSPY1 – *TTY1* + 2. The deletions thus have the order alphoid – TSPY2 + 1 – *TTY1*, having lost *AMELY*, *PRKY*, and *TTY2*. This is illustrated in figure 6.14, below.

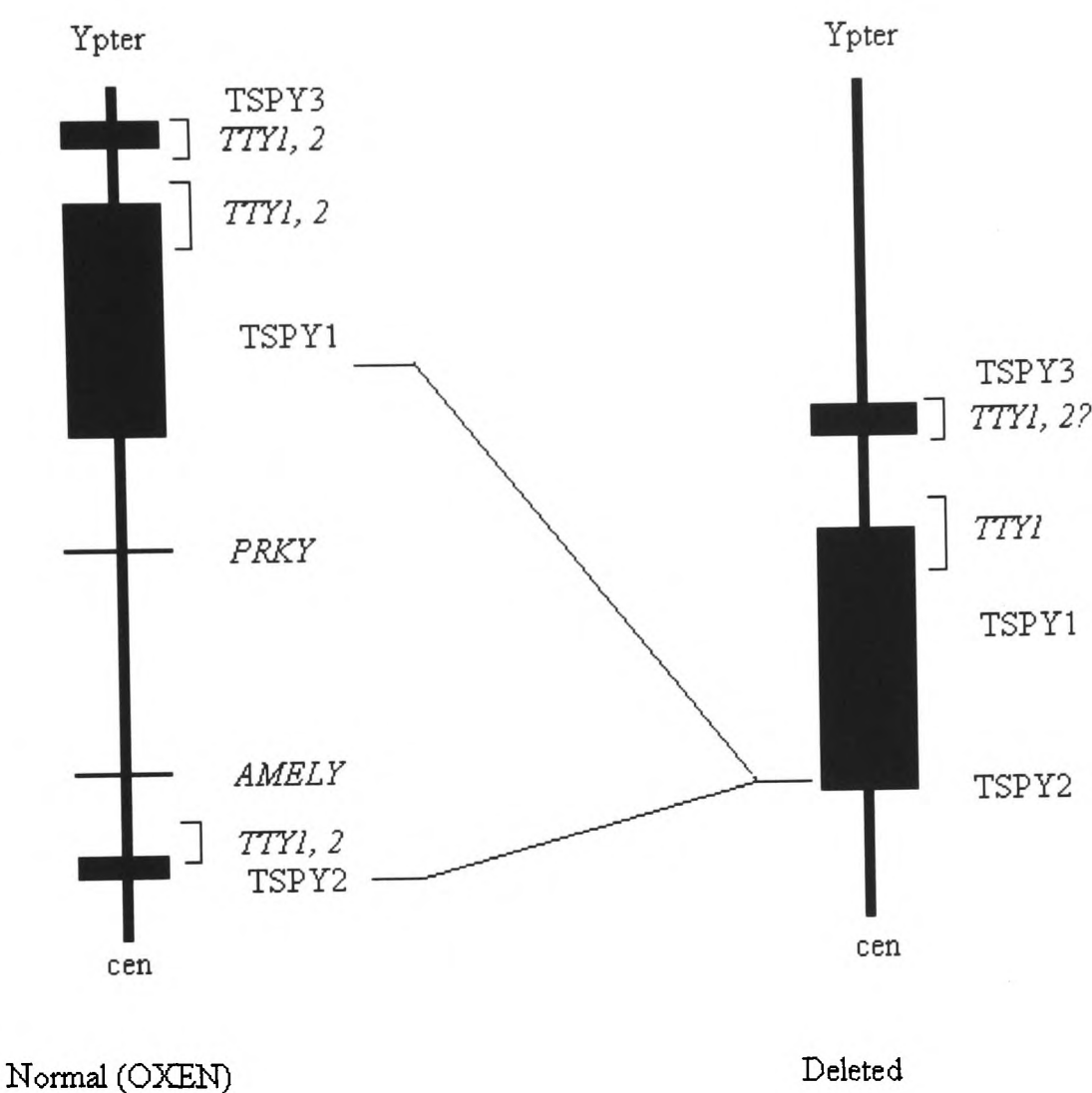


Figure 6.14: The locations of probes on the normal and deleted forms of the Y chromosome.

This allows us to separate the two *TTY* genes, as *TTY2* has been lost in the recombination, while *TTY1* has not, giving the original OXEN-type order as alphoid – TSPY2 – *TTY1* + 2 – *AMELY* – *PRKY* – TSPY1 (with *TTY2* sequences as part of the array) – *TTY1*. The *PRY* gene, on the other hand, was not seen to be lost; which is consistent with our localisation on the long arm, and the short arm locus failing to be detected. Lahn & Page also localise a second copy / cluster of *TTY2* sequences in interval 6C on Yq. Possibly the primers used here do not amplify these sequences, so that the *TTY2* sequence appears to be absent from the deleted chromosomes. An additional complication is provided by the apparent absence of *TTY2* sequences around the TSPY3 block on the deleted chromosomes – this apparent absence may

reflect difficulty in amplification from genomic DNA, as the sequences are present in YACs from the TSPY3 region. Alternatively, more extensive rearrangements of the deleted chromosomes cannot be ruled out, resulting in the loss of the *TTY2* sequences. These results suggest that *AMELY*, *PRKY*, and *TTY2* are not required for normal male development, as the deleted individuals were, as far as is known, normal.

The second set of experiments carried out on deleted Y chromosomes was concerned with localising a gene involved in Turner syndrome. A recent report (Tzancheva et al., in preparation) described two patients with ring Y chromosomes, one of whom (TS) had Turner syndrome, while the other (DM) did not. An analysis of the ring chromosomes showed that both individuals had all of the short arm loci examined, from *SRY* (judging by their phenotype) to *DYZ3*, while TS lacked all long arm sequences tested, and DM had long arm loci only as far as the *RBF5* probe, but lacked the *GMGY6* sequence.

This suggests that some gene in the region present in DM but absent in TS has some function as an anti-Turner syndrome gene. Accordingly, DNA samples from both these individuals were tested for the presence or absence of the *DFFRY* gene, as this is the only identified gene from the appropriate region. The result is shown in figure 6.15 below. Other loci were tested (*DBY*, *UTY*, *BPY1* and *TB4Y*), and confirmed that this more distal region was absent from both individuals.

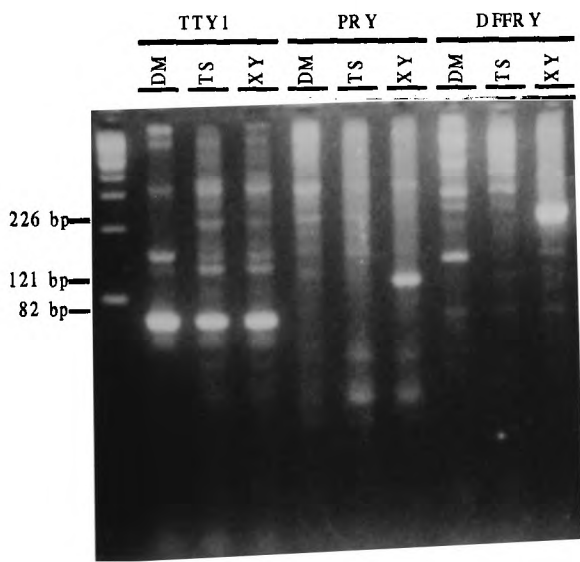


Figure 6.15: PCR of DM and TS for the *DFFRY* gene.

This shows that both individuals, the Turner patient and the non-Turner patient, lacked the *DFFRY* gene. Recalling the previous section, it will be remembered that the precise location of *DFFRY* relative to RBF5 was undefined, insofar as it was unknown whether the gene was proximal or distal to RBF5. The present result suggests that the gene is distal to RBF5, hence both individuals have lost the gene. (Of course, an alternative explanation is that *DFFRY* is actually proximal to RBF5, but patient DM has undergone further rearrangements of the Y chromosome in addition to a simple deletion). Whatever the true explanation, this result discounts the possibility that *DFFRY* is the anti-Turner syndrome gene (this is supported by the work of Brown et al. [1998], who suggest that *DFFRY* is the AZFa gene, rather than a Turner gene). Therefore, there must be one or more genes on the long arm remaining proximal to RBF5 which are involved in Turner syndrome. For instance, the map shows a number of restriction site clusters, which are linked with the presence of genes, such as the *BsiWI* / *BssHII* / *MluI* / *SacII* cluster distal to the probe 109e, or the *BssHII* / *MluI* / *SacII* cluster proximal to 219e. Such conclusions suggest a direction in which further research may be taken.

6.3: A study of sex-reversed individuals.

The final investigation into functional aspects of these genes was carried out with sex-reversed individuals. It will be recalled from section 6.1 that the *PRKY* gene has been observed to act as a hotspot for illegitimate recombination between the X and the Y chromosomes (Klink et al., 1995). This recombination leads to XX males, who have the *SRY* region transferred onto the X chromosome, and XY females, who have lost the equivalent region from their Y chromosome. This type of sex reversal is rare amongst XY females, the majority of whom possess the *SRY* region, and have mutations elsewhere in the genome. However, it is reported to be the most common genetic cause of XX males, accounting for one third of all XX males (Schiebel et al., 1997).

The *PRKY* gene lies on the short arm of the Y chromosome, and is included in the inverted region described in section 5.3. Since the *PRKY* recombination relies on both the X and the Y regions being homologous, it would seem likely that the orientation of the inverted region on the Y chromosome affects the occurrence of the recombination. That is to say, only those Y chromosomes with a particular orientation

of the *PRKY* gene would be susceptible to giving rise to this class of sex-reversed individuals. In order to test this hypothesis, it was decided to analyse as many sex-reversed individuals with the *PRKY* crossover as possible.

In the study of Y chromosomal loci mentioned above (Pandya, 1998), a Y haplotype tree was constructed showing 19 different Y chromosomal haplogroups. The tree for the European populations is reproduced below, in figure 6.16.

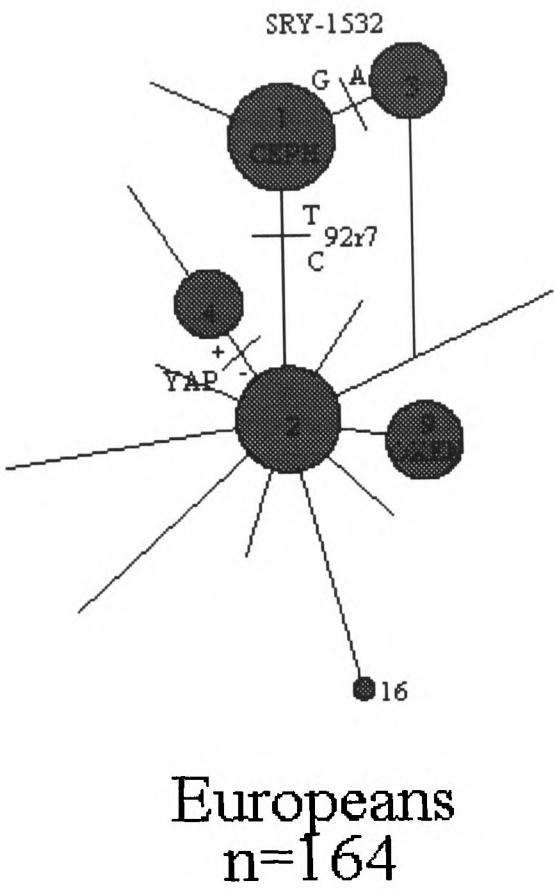


Figure 6.16: The European Y haplogroup tree (from Pandya 1998).

Each circle on the graph represents one haplogroup, the size of the circle being proportional to the number of individuals in each haplogroup. Each edge of the graph signifies a single polymorphism joining the haplogroups; those relevant to the current work have been marked on the graph. The haplogroups are numbered in accordance with Pandya (1998). Only 6 of the 19 haplogroups found worldwide were present in a sample of 164 Europeans, with the relative frequencies of each haplotype given in table 6.1 below.

Table 6.1: Relative frequencies of each of the European haplogroups (Pandya, 1998)		
Haplogroup	Number	Frequency
1	48	0.30
2	46	0.28
3	24	0.15
4	19	0.12
9	25	0.15
16	2	0.01

The two forms of the inversion have been placed on this tree, with the Boleth cell line (being of the CEPH type) falling in haplogroup 1, and the OXEN cell line being a member of haplogroup 9. The inversion must therefore lie either between haplogroups 1 and 2, or between 2 and 9. A location between haplogroups 1 and 2 was considered more likely, as an additional polymorphism (M911) occurring between these two groups is located on the proximal TSPY block (TSPY2). It is thought that this M911 polymorphism may be diagnostic of the inversion, showing different forms depending on the orientation of the chromosome – thus, all members of haplogroups 2 and 9 are of the OXEN type, while members of haplogroups 1 and 3 are of the CEPH type.

PCR-based polymorphisms were used to identify the haplogroups to which the sex-reversed samples belonged. These polymorphisms were: 92r7, which has a C → T transition which destroys a *Hind*III site, occurring at the boundary between haplogroups 2 and 1. The 92r7 PCR amplifies multiple (at least four) loci, only one of which is polymorphic. The non-polymorphic loci are all of the T type, while the other locus may be either the T or the C form. The *PRKY* crossover occurs between these loci, so the sex-reversed individuals only have a single 92r7 locus, which is the polymorphic one. SRY-1532, which has a G → A mutation destroying a *Dra*III site, occurs between haplogroups 1 and 3; and YAP, an insertion of an Alu sequence at the boundary between haplogroups 2 and 4. Taken together, these polymorphisms could place the *PRKY* crossovers on the haplotype tree. The physical locations of these polymorphisms on the chromosome are shown in figure 6.17 below. It will be observed that the polymorphisms can only be tested on some crossover samples: SRY-1532 and 50f2/B will be present only on the XX male samples; while XY female samples will only possess the 92r7 and YAP loci. Fathers of subjects, where available, will have all loci on the intact Y chromosome, and so will be very

informative; however, it is rare that paternal samples can be obtained of sex-reversed subjects.

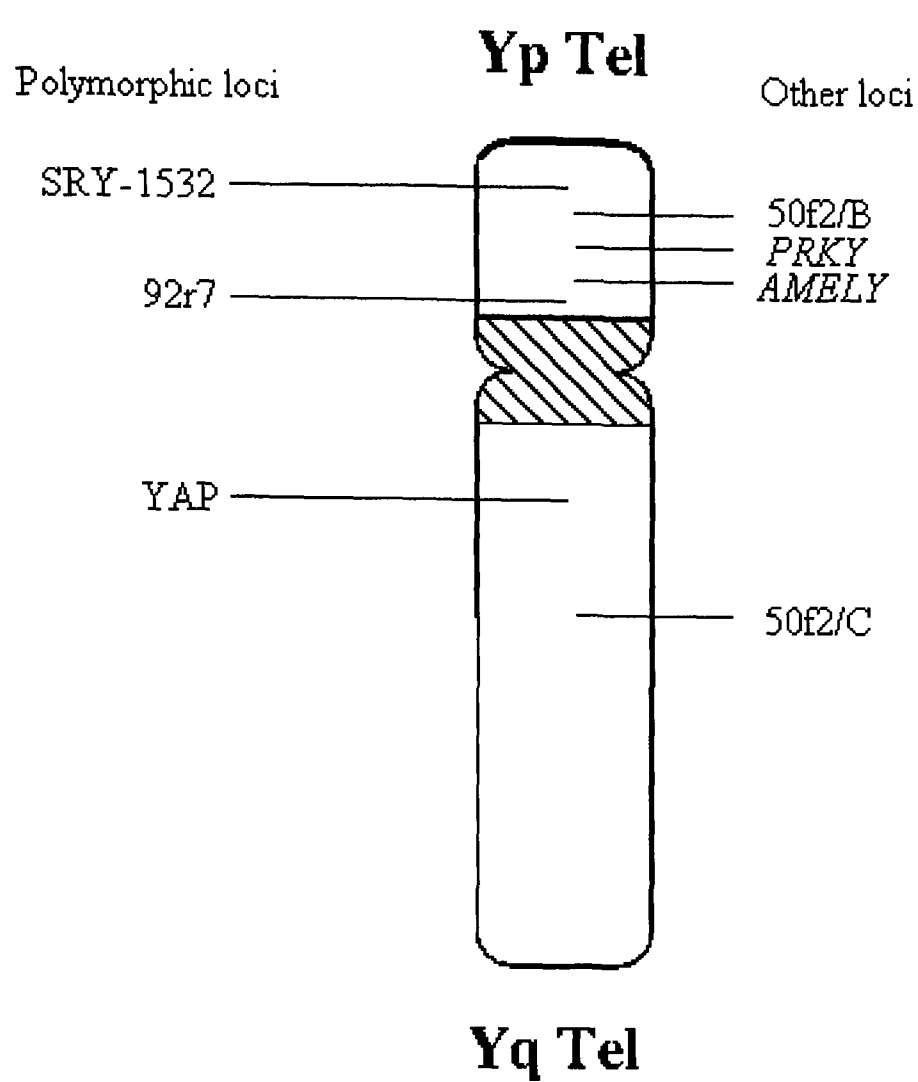


Figure 6.17: Locations of the polymorphic loci tested on the normal Y chromosome.

70 DNA samples from XX male and XY female subjects, and some normal relatives were obtained from Gudrun Rappold (Heidelberg), Ken McElreavey (Paris), Mark Jobling (Leicester), and Nabeel Affara (Cambridge). Some had been characterised in detail and were known to have breakpoints in the *PRKY* region. Less information was available about others, and, where necessary, the presence or absence of the loci 50f2/B and *AMELY* was determined. Breakpoints between these loci were considered to lie in the *PRKY* region. 31 XX males and 5 XY females were determined to have *PRKY* breakpoints. These samples were then tested with SRY-1532 (males), or 92r7 and YAP (females). Relatives of the subjects, when present, could be tested with all markers. The PCRs are shown in figures 6.18 – 6.21 below; and the samples and results listed in table 6.2. Only those samples which were tested were scored; a blank entry in the table denotes that the sample was not tested for that locus.

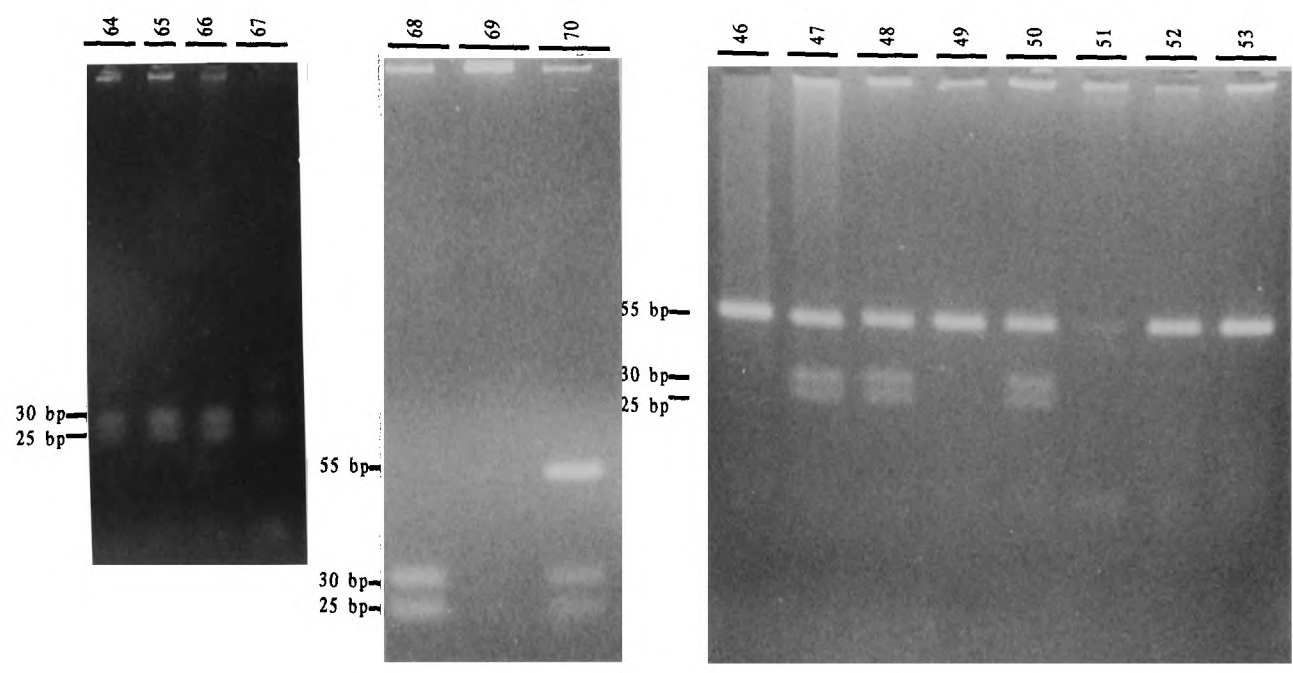


Figure 6.18: PCR of XY females with 92r7, digested with *HindIII*. The C allele of the polymorphism has a restriction site, giving products of 25 and 30 bp. The T allele lacks the restriction site, and yields only a 55 bp product. The numbering of the lanes refers to the numbering given in table 6.2.

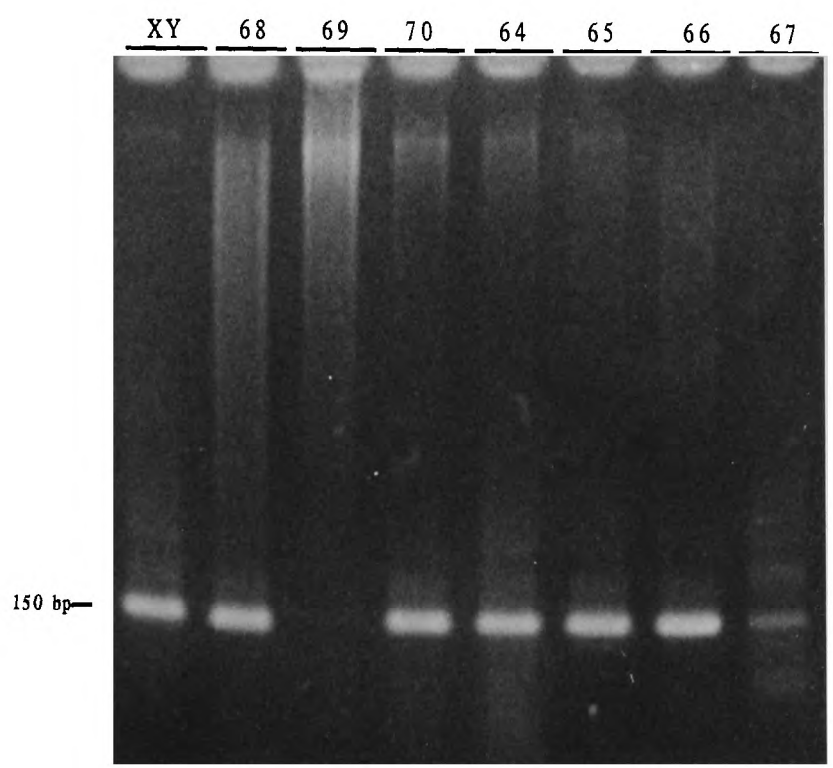


Figure 6.19: PCR of XY females with YAP. The YAP- form yields a 150 bp product, while the YAP+ form gives a product of 455 bp. The numbering of the lanes refers to the sample numbers given in table 6.2.

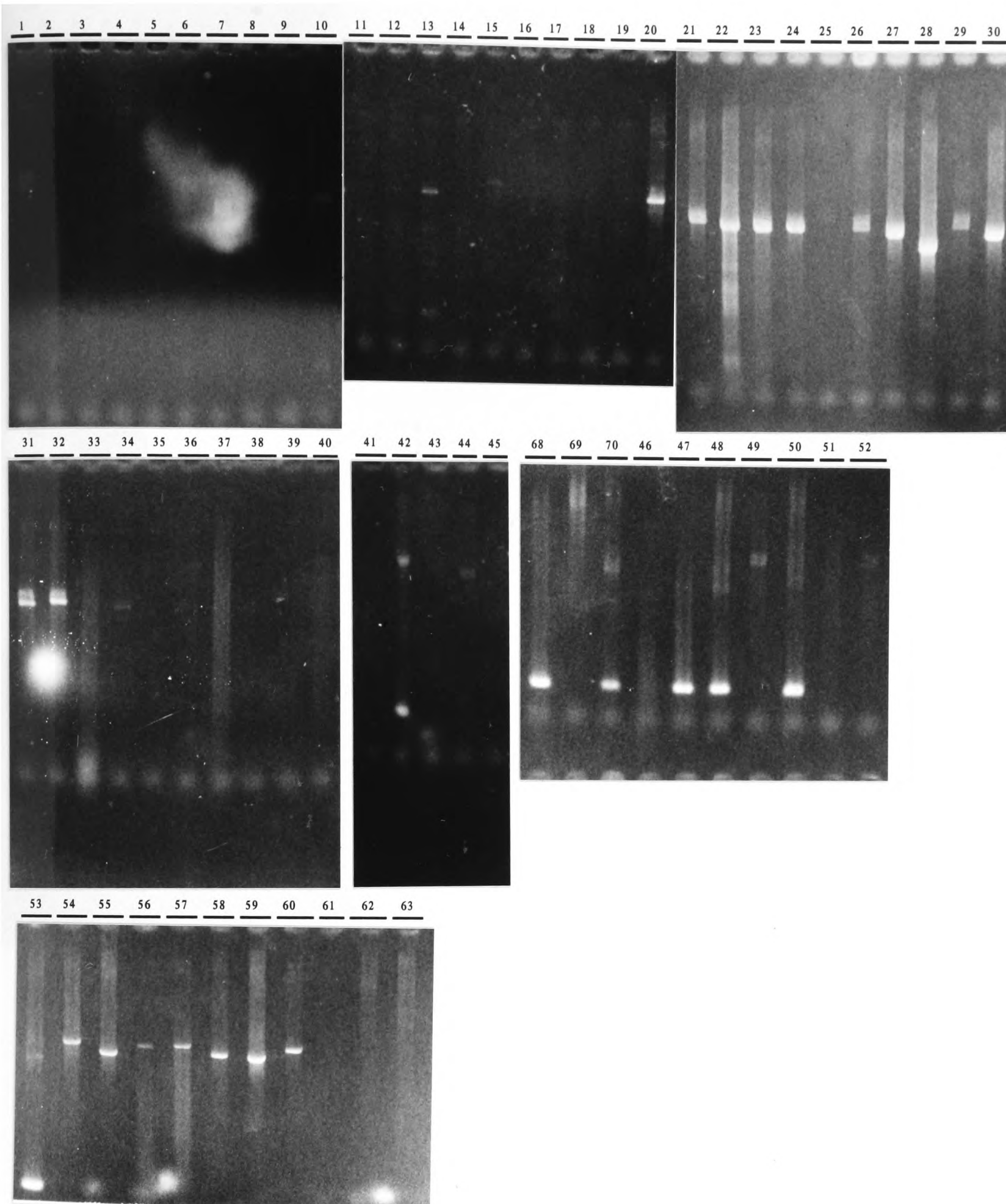


Figure 6.20: PCR of XX males with 50f2. The size of the B locus is polymorphic, but is fairly large. The C locus is smaller. The lane numbering refers to table 6.2. The results from samples 6-8, 25, 36, and 47 were confirmed in other experiments. Sample 42 was here switched for normal male DNA; and retested subsequently.

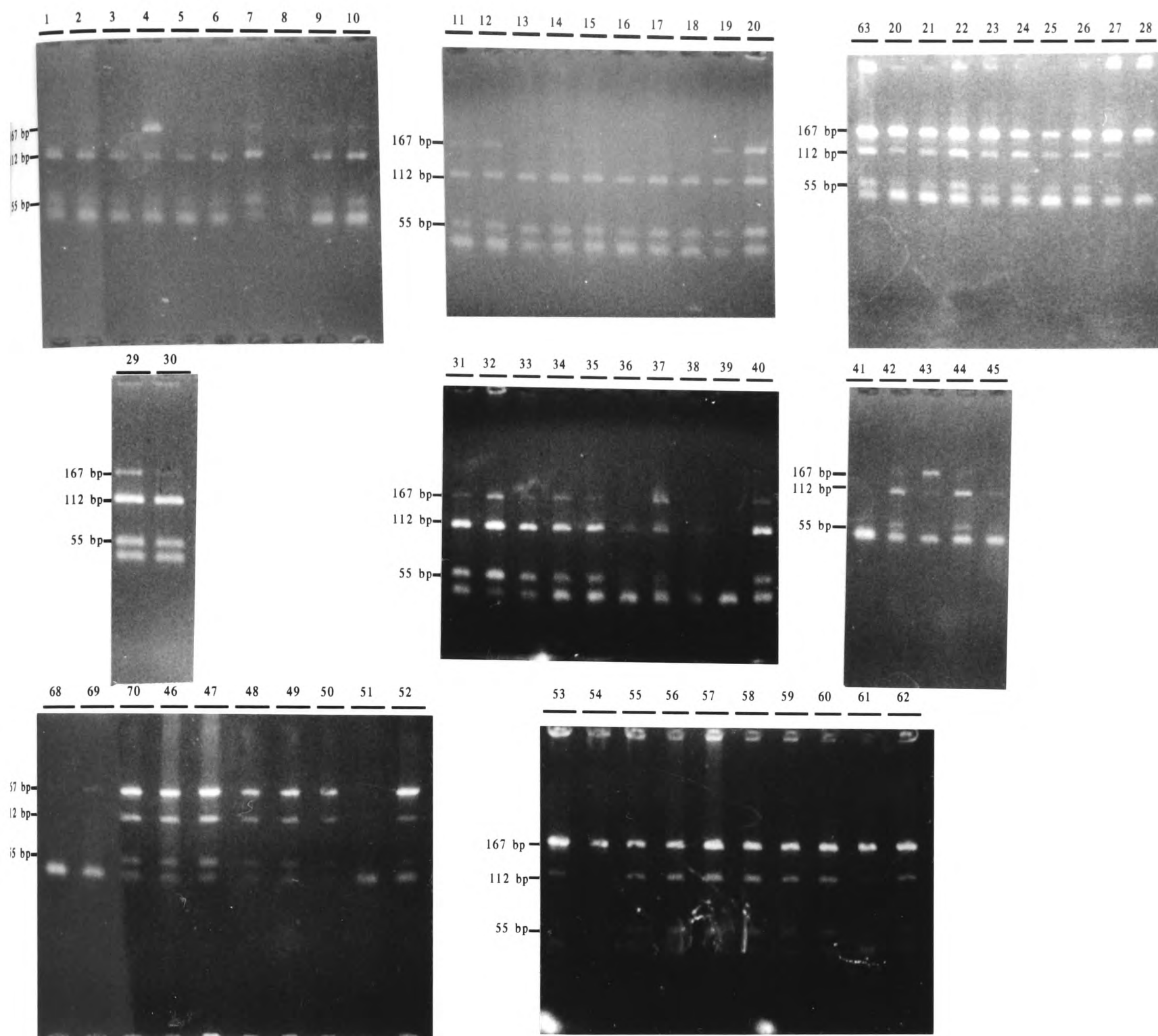


Figure 6.21: PCR of XX males with SRY-1532, digested with *DraIII*. The G locus has a restriction site, and gives products of 55 and 112 bp; while the site is lacked by the A locus, which gives a product of 167 bp. Digestion is often incomplete, so the G allele is scored from the presence of the 55 bp and 112 bp bands, rather than the absence of the 167 bp band. Samples 28, 38, 39, 41, 54 and 61 were repeated in order to generate the results shown in table 6.2

Table 6.2: Summary of the sex-reversed samples tested, and the results.							
Name	Number	Source	Type	50f2	SRY	92r7	YAP
1	1	KM	XX male	B	G		
2	2	KM	XX male	-	G		
3	3	KM	XX male	-	G		
4	4	KM	XX male	-	G		
5	5	KM	XX male	-	G		
6	6	KM	XX male	B	G		
7	7	KM	XX male	B	G		
8	8	KM	XX male	B	G		
9	9	KM	XX male	-	G		
10	10	KM	XX male	B	G		
11	11	KM	XX male	-	G		
12	12	KM	XX male	-	G		
13	13	KM	XX male	B	G		
14	14	KM	XX male	-	G		
15	15	KM	XX male	B	G		
16	16	KM	XX male	-	G		
17	17	KM	XX male	-	G		
18	18	KM	XX male	-	G		
19	19	KM	XX male	-	G		
Kruse	20	MJ	XX male	B	G		
McBean	21	MJ	XX male	B	G		
CHA	22	MJ	XX male	B	G		
WLA	23	MJ	XX male	B	G		
Critelli	24	MJ	XX male	B	G		
LGL105	25	GR	XX male	B	G		
GA001	26	GR	XX male	B	G		
CHM347	27	GR	XX male	B	G		
WHT1856	28	GR	XX male	B	G		
WHT2127	29	GR	XX male	B	G		
WHT950	30	GR	XX male	B	G		
WHT1123	31	GR	XX male	B	G		
WHT1209	32	GR	XX male	B	G		
693	33	GR	XX male	-	G		
RA	34	GR	XX male	B	G		
ZO	35	GR	XX male	-	G		
ED	36	GR	XX male	B	G		
547	37	GR	XX male	-	G		
Ti	38	GR	XX male	-	G		
548/11	39	GR	XX male	-	G		
462	40	GR	XX male	-	G		
HH	41	GR	XX male	-	A		
JY	42	GR	XX male	-	G		
MK	43	GR	XX male	-	G		
101	44	GR	XX male	B	G		
CF	45	GR	XX male	-	G		

JMcC	46	NA	XX male		G	-	
AMcC	47	NA	XY brother of JMcC	B + C	G	C	
JoMcC	48	NA	XY father of JMcC	B + C	G	C	
AG	49	NA	XX male	B	G	-	
AG2	50	NA	XY father of AG	B + C	G	C	
EG	51	NA	XX mother of AG	-	-	-	
DP	52	NA	XX male (but contains AMELY)	B	G	T or -	
WP	53	NA	XY father of DP	B + C	G	T	
JT	54	NA	XX male	B	G		
TA	55	NA	XX male	B	G		
GA	56	NA	XX male	B	G		
WB	57	NA	XX male	B	G		
RH	58	NA	XX male	B	G		
DR	59	NA	XX male	B	G		
PW	60	NA	XX male	B	G		
SB	61	NA	XX male	-	G		
DW	62	NA	XX male	-	G		
KS	63	NA	XX male	-	G		
WHT745	64	GR	XY female			C	YAP-
WHT1615	65	GR	XY female			C	YAP-
651	66	GR	XY female			C	YAP-
WHT1327	67	GR	XY female			C	YAP-
AM	68	NA	XY female	C	-	C	YAP-
CM	69	NA	XX mother of AM	-	-	-	-
JM	70	NA	XY father of AM	B + C	G	C	YAP-

Overall, 70 samples were tested. Dealing with the XY females first of all, there were 5 of these individuals. In addition, we obtained two relatives of AM, one of the XY females: CM, the mother, and JM, the father. It will be seen that the fathers of the *PRKY* class of XY females will have the same chromosomal organisation as the Y part of XX males; and *vice versa*.

All five of the XY females were of the form YAP-, and 92r7 / C. This shows that the Y chromosome of these individuals belongs to haplogroups 2, 9 or 16. (That is to say, all these *PRKY* exchanges occur on a background of the OXEN form of the

inversion.) From the data given of the frequencies of these haplotypes, we can calculate the probability of this grouping. We cannot predict the haplotype of the first individual, so this has a probability of 1. However, we then predict that all additional individuals should belong to the same set of haplogroups. The probability that a random European Y chromosome belongs to groups 2, 9 or 16 is 0.44. The probability that the next four samples will all belong to this grouping is $(0.44)^4$, or approximately 0.037. This is within the limits of the standard 95% confidence interval, and so we can conclude that the XY female *PRKY* crossover occurs on chromosomes of the OXEN type.

The XX males were more complicated to study. Initially, we obtained 58 XX male samples, along with 5 relatives of some of the individuals. As mentioned above, the fathers of the *PRKY* class of XX males will have the same chromosomal organisation as the XY *PRKY* females. The 58 samples were first tested with 50f2, for the presence of the B locus – any individuals lacking this were rearranged more distally than the *PRKY* gene, and so were discounted from the analysis. This left 32 individuals with the appropriate breakpoint. One of these 32 (sample number 52, DP), however, was known to have the *AMELY* locus present (Sargent et al., 1996). As the organisation of the chromosome is Ypter – *SRY* – 50f2/B – *PRKY* – *AMELY*, this sample again had to be discounted as having a different rearrangement from the remaining samples (as did the father of DP, WP). This left 31 samples, which, when tested with the SRY-1532 polymorphism were all found to possess the G allele.

The frequency of the G allele in the European population is, from table 6.1, 0.85. Using the same argument as for the XY females above, the chance of all these samples coming from this group is $(0.85)^{30}$, or 0.006. Again, this is well below the level required for a 95% confidence interval. This demonstrates that the class of XX males with a *PRKY* breakpoint forms a limited subset of the total Y chromosome population. Unfortunately, this analysis does not exclude these males from having the CEPH form of the inversion, group 1 being of this type. In order to exclude this possibility, the fathers of the XX males have to be considered. As only 2 of the fathers were in this analysis, no statistically valid conclusions can be drawn from the fact that both fathers had chromosomes of the OXEN type. It may be possible to combine the data from the fathers of the XX males, and the XY females, as both these groups

represent the same type of Y chromosome; in this case, the probability that all seven samples are of the same haplogroup is $(0.44)^6$, based on the results of the 92r7 polymorphism; this gives a probability of 0.0073 that this result would occur by chance.

So, we see that all these *PRKY* sex-reversed individuals form a limited subset of the available Y chromosome types; the XY females all being of the OXEN form; while the XX males cannot be assigned to a single type because they lack the region of the Y chromosome required for testing, although they still do not occur on all Y chromosome backgrounds. Is this evolutionarily significant? Presumably, it is a disadvantage to be prone to having sex-reversed offspring, as these are almost invariably sterile. Even if the frequency of occurrence of sex reversal is low, this may still give a selective advantage to those haplotypes which will not give rise to sex reversal.

Further work in this area has to be carried out – in particular, more examples of fathers of XX males need to be obtained in order to confirm or deny the hypothesis that all *PRKY* crossovers only occur on the OXEN form of the inversion.

Chapter 7: Discussion

The aim of this thesis has been to assemble a physical map of the pericentromeric region of the human Y chromosome, and to study the organisation of the chromosome. In this chapter, I shall review the results obtained, and give directions in which future work may be carried out.

7.1: Physical YAC mapping

The initial part of this work was to extend the YAC contigs previously assembled from this region, and to attempt to close the gaps between the contigs. The study began with six separate contigs, three from the short arm, and three from the long arm. Work commenced with screening the CEPH mega-YAC library with a selection of probes from the extremities of each of these contigs.

7.1.1: 14A3C and GMGY6

Seven YACs were mapped onto these two contigs, one of which, 294g5, was found to close the gap between the two. It was already known from PFGE mapping of genomic DNA that the two contigs were very close, a fact which was confirmed by the mapping of this YAC, only some 360 kb in length, and which showed that the gap between the contigs was only 50 kb. The final unified contig covers 2300 kb of DNA, and is well represented in both CEPH and OXEN library YACs.

Future work may be undertaken in order to extend further the coverage of the contig. No great difficulties are foreseen in extending the map in the Yqter direction; this is merely a matter of time and resources to devote to the problem. This area is, however, outwith the region under study in this present work, and so no attempt was made to extend the map in this direction. Furthermore, the more distal part of Yq has already been fairly intensively studied by other groups (see, for example, Kirsch et al., 1996), due largely to the fact that several putative genes or genetic functions have been identified in this region (such as *TB4Y*, *AZF*, *GXY*, and others – refer back to chapter 1).

Extending the 14A3C / GMGY6 contig towards the centromere may prove more difficult. 14A3C is the most proximal probe from this region, and screening of the

OXEN and CEPH mega-YAC library did not provide any further YACs to extend the contig. From PFGE analysis of genomic DNA, the contig has been shown to lie at most 1 Mb from the more proximal 68 bp satellite contig; however, it will be necessary to screen additional libraries to extend the map in this direction.

7.1.2: The 68 bp satellite contig

As mentioned throughout chapter 3, this contig has proved extremely difficult to extend. The problems have two causes: first, the probes which do lie in the contig are satellite and other repeated sequences which are present throughout the genome; thus, any library screen (YACs or other cloning methods) would produce an unmanageable number of clones to be studied. Furthermore, these probes are not Y-specific sequences; unlike other repeated probes which are of use in mapping studies, there is often no way of distinguishing the Y-chromosomal copies of the satellites from the other genomic copies. So all clones selected in mapping this region would have to be studied in detail to confirm a Y-chromosomal origin.

The second, and more serious, cause of the mapping problems is due to the under-representation of other probes in libraries. The 64a probe is a useful probe detecting only two loci; however, as mentioned previously, does not select any Y chromosomal YACs from the CEPH mega-YAC or the OXEN libraries. Two PACs were identified which linked the 64a probe to the satellite sequences, showing that such an approach to completing the map would be feasible.

It seems likely that the absence of any 64a-containing YACs is due to some inherent instability or growth inhibition of this region in the YAC vector, presumably a consequence of the nearby satellite repeats. In order to extend the contig in either direction, toward the 14A3C or alphoid contigs, other methods of mapping will have to be investigated. Closure of the alphoid gap may prove to be easier, given the existence of the non-satellite 64a probe; the most distal probe toward the 14A3C side is KFC43, another highly repeated sequence. Of course, beginning to close the gap would probably be easier from the 14A3C contig itself; however, the most proximal YAC from this contig, yOX 85, contains some fragment of the 5 bp satellite sequence (C. Tyler-Smith, pers. comm.), so suggesting that extending the map from this side would also suffer from the problems of mapping in satellite regions.

In addition to these difficulties, the scale of the task required to close the gaps must not be under-emphasised. In switching to non-YAC cloning systems, the major advantage of the large size of YACs is given up. With the gaps being estimated at 1 Mb (to 14A3C) and 700 kb (to alphoid), the relatively smaller insert size of alternative systems mean that much effort will have to be put in to closing the gaps. Further, a c.400 kb array of the 5 bp satellite has been identified adjacent to the alphoid DNA, and it is not clear how such a repetitive array could be mapped using clones of 150 – 200 kb.

Of course, the gaps have already been closed in genomic DNA, and so the region is not entirely absent from our current knowledge about the chromosome. However, it would be of utility to have the entire proximal long arm region covered in cloned DNA, as this would facilitate analysis of the fine details of the region; unfortunately, this is unlikely to happen, thanks to the very large effort required to close a small gap.

7.1.3: The alphoid – TSPY contig

The major piece of work involved in the mapping of this contig was the analysis and mapping of the YAC 887g12, which eventually closed the gap between the alphoid and TSPY contigs. This YAC proved particularly difficult to map thanks to the fact that it was both extremely large and chimaeric, with both ends containing non-Y chromosomal DNA. It was known that it did contain sequences from both the alphoid and TSPY contigs, and so the effort was invested in assembling the map. Initial mapping of the intact YAC proved fruitless, since the large size (over 2 Mb) meant that the internal restriction sites could not be visualised from partial mapping; and the left arm did not appear to fit with the map of the TSPY contig, so suggesting its chimaeric nature.

Accordingly, it was decided to fragment the YAC, in order both to remove some of the chimaeric sequences, and to reduce the Y-specific region of the YAC to a more easily studied size. Upon fragmentation, a very low yield of transformants was obtained. Further analysis revealed that this was due to the presence of two copies of the original YAC in each cell, thus ‘true’ transformants, detected by both positive and negative selection procedures, would only appear when both copies of the original

YAC were fragmented, or the second copy was lost or mutated. Thus, after obtaining three transformants of a suitable size, these clones had to be purified further, by growth without selection until one of the copies of the YAC was lost.

The two smallest of these clones were mapped, and 887Δ4 was seen to link the alphoid and TSPY contigs. This in itself suggested that a proposed inversion on the short arm may exist, as the standard deletion maps and YAC contigs had previously suggested that a link should exist between the alphoid and *AMELY* contigs. If further mapping work were to be carried out in this region with additional OXEN YACs, it would be expected that a divergence between the CEPH and OXEN libraries will be revealed at some point in the region covered by 887Δ4.

With this YAC mapped, and a number of other YACs placed on the unified alphoid – TSPY contig, the region covered now extends over approximately 2 Mb (this is only approximate as the sizes of both the TSPY1 and alphoid arrays is variable between individuals). In total, five new YACs were placed on these contigs, all from the CEPH library. One observation that may be made from the map is that seven YACs terminate just inside the major alphoid array, reflecting again the instability of certain repetitive sequences in YACs.

Closure of the gaps remaining from this contig to the 68 bp satellite and *AMELY* contigs again presents difficulties. As with the 14A3C contig, the sequence closest to the 68 bp satellite contig is a block of 5 bp satellite sequence; thus the gap may well prove very difficult to close from this end, as well as being troublesome from the 68 bp satellite side of the gap. However, it bears repeating here that this particular gap has been covered in genomic DNA.

Attempting to close the gap from TSPY1 to the *AMELY* contig could, again, prove problematic. The distance between the TSPY1 block and the most proximal region of the *AMELY* contig is unknown, due to the sparsity of suitable probes for use on genomic DNA – the LLY22g probe is moderately repeated, with at least 7 copies in the genome; while 758e on the *AMELY* side appears to be highly repeated. From the library screenings, several TSPY1-containing YACs have been found, although the

available evidence suggests that none of these YACs cover the appropriate part of the TSPY contig, so there are no new candidates for generating novel Y-specific probes to close the gap, which would make screening the library far simpler. The rightmost end of the YAC 759g2*, which extends past the LLY22g/F sequence, was amplified by means of Alu-vector PCR, but the sequence so obtained was too repetitive to be useful as a probe. Thus, there are no simple probes from this region which can be used to close the gap. So, in this case, as with the 68 bp satellite gaps, there seem to be no obvious reasons why the gaps could not be closed with time, although a disproportionate amount of effort would be required.

7.1.4: The *AMELY* contig

This is the largest single contig of the map, covering some 4 Mb of sequence, largely from the CEPH library, although there is extensive coverage in OXEN YACs as well. During the course of this work, the contig was extended by some 1.5 Mb towards Ypter, using solely CEPH YACs, causing the map to extend some way into the main X-Y homologous region on the short arm. This part of the contig marks the extreme limit of the coverage of the map, as the more distal sequences have already been mapped quite extensively in YACs (see, for instance, Whitfield et al., 1995 [for the PAR and *SRY* regions], and Sargent et al., 1996 [for the X-Y homologous blocks]). No problems were experienced in extending the contig this far, once the X-Y homologous nature of the region was known; and no great difficulties would be envisaged were the mapping coverage to be extended towards the short arm telomere, given that the majority of X-Y homologous sequences from this region have chromosome-specific features.

The proximal end of the *AMELY* contig extends toward the TSPY contig. This appears to be a more difficult region to map. At the outset of this work, a substantial part of the region was covered in only a single YAC, 758g1, terminating in the probe 758e. This probe is fairly poor if used for screening YAC libraries, thanks to the number of similar sequences it detects throughout the genome. Making use of the proximity of the 50f2A and 50f2B loci to this sequence, however, did allow two YACs to be selected from the CEPH and OXEN YAC libraries, and placed on the map: yOX 81 and 931-ura-6. This last YAC was particularly difficult to map, as it

proved very difficult to obtain clones with a single YAC present. As the 931-ura-6 clone contained only a single YAC, and this YAC did not hybridise to the standard right vector arm probe, it seems that this was due to the clone containing two YACs which had lost one of the genes used for selection of YACs – hence, the yeast could only survive under selection if both YACs were present. Purification of a single YAC was eventually obtained by selecting only on the left vector arm, so allowing the non-Y YAC to be lost.

The proximal part of this contig, from 758e itself to at least as far as *AMELY*, is thought to be involved in the short arm inversion referred to in section 7.1.3. And, again, as with the YAC 887g12, it would be expected that there will be differences in the maps derived from the CEPH and the OXEN libraries, marking the breakpoint of the inversion. Obviously, these differences have not yet been discovered, thanks in part to the low level of coverage of the more distal parts of this contig in OXEN YACs. If these YACs were to be investigated in some detail, and more OXEN YACs covering this part of the contig were to be mapped, then the location of the breakpoint may become more apparent. This is discussed further below.

So, to summarise the progress in assembling the contig maps, it will be seen that substantial progress has been made toward the ideal of having a single contig stretching from the proximal short arm, through the centromere, to the proximal long arm. The original six contigs have now been joined to give 4 contigs, with one gap on the short arm between the *AMELY* and TSPY contigs; and the remaining two gaps separating the repetitive 68 bp satellite contig from the alphoid and 14A3C contigs on the long arm. The orientation of each of the contigs has been established, so making further construction of the map simpler.

Currently the YAC maps cover 6.5 Mb on the short arm, with the one gap being of unknown size; and 2.6 Mb on the long arm, with the gaps in this region being a maximum of 1.7 Mb in total.

Prospects for extending the map still further are low, but not zero. The major problem in extending the contigs would be the lack of suitable probes for screening clone libraries, the most extreme points of each of the contigs being composed of satellite or

other repetitive DNA, with some of these sequences seeming to be under-represented in the YAC libraries. So, in order to continue extending the map, it would apparently be most profitable to begin making use of alternative cloning mechanisms, such as the PAC library used briefly in chapter 3, or whichever other libraries are readily available. The great disadvantage of this approach is, of course, the relatively smaller size of the insert in other cloning systems – a not untypical CEPH mega-YAC will be 1000 kb in size; the PAC mapped in chapter 3 was only 200 kb in size. Thus, even allowing for the greater incidence of chimaerism among YACs, a far greater effort must be invested in this part of the map than in the YAC mapping. However, there would seem to be no obvious reasons why such an attempt at mapping would fail, were it to be undertaken.

7.2: Sequence features of the Y chromosome

Armed with the finalised version of the chromosomal map, several aspects of the sequence organisation of the Y were investigated. Initially the distal part of the *AMELY* contig was identified as being part of the large Yp / Xq homologous region, by virtue of its probe content – GMGXY5, 9, and 12. This also established a definite link between this map and the previously published map of this X-Y homologous region, and showed the extent of this coverage in the present map, as the next probe, GMGXY4, was not present in any of the YACs placed on the *AMELY* contig.

Secondly, a duplication of the alphoid contig AA15e – 52d/C – TSPY block was suspected to exist on the *AMELY* contig, thanks to the unexpected presence of these three probes in one of the earliest YACs to be mapped to this contig, 752h1. Detailed investigation of this YAC and another mapped to the same area (894f6) showed the location of this duplication to lie between the two X-Y homologous regions on the short arm, the one described in the previous paragraph, the other defined by the probes GMGXY8 and pDP34 (Figure 6.10 may be referred to here to refresh the memory).

The existence of this duplication seemed to be of little import, other than another example of the type of repeated sequences which make the Y chromosome so difficult to map, until the discovery of an inversion occurring on the short arm. This inversion occurs between the Boleth and the OXEN cell lines, the source chromosomes for the

CEPH and OXEN YAC libraries, and is indicated by the probe ordering alphoid – *AMELY* – TSPY1 in the OXEN type, and alphoid – TSPY1 – *AMELY* in the CEPH type. The limits of the inverted region have not been precisely defined, nor the mechanism whereby the inversion occurs.

The existence of the alphoid / *AMELY* contig duplication, and the determination of the orientation of each of the contigs, suggests one means by which the inversion could take place. The duplicated regions described are in opposite orientations, so forming an inverted repeat region flanking the block of inverted sequence. The inverted repeats may undergo homologous recombination within a chromosome, so resulting in the intervening sequence changing orientation. A similar mechanism involving inverted repeats has been proposed to occur on the X chromosome emerin gene (Small & Warren 1998), although on a much smaller region of sequence. In the current case, it is likely that at least 4 Mb of sequence is involved in the inversion.

However, the proximal end of the inversion must be within, or distal to, the TSPY2 array, since single YACs from each type of chromosome link the alphoid array directly to TSPY2 (the CEPH YAC 62C12, and the OXEN YAC yOX27, for example). An attractive hypothesis is that the inversion occurs between TSPY2 and TSPY3 sequences, or the adjacent duplicated region – this is illustrated in figure 7.1 below.

This hypothesis must be rejected because it predicts that *AMELY* should be closely linked to AMY8e only in the CEPH type chromosome. In fact, AMY8e is linked to *AMELY* by uninterrupted YACs in both libraries (such as 757h5 from the CEPH library, and yOX25 from the OXEN library). The model thus described could only be true if AMY8e was duplicated in genomic DNA, but always deleted from YACs close to the alphoid DNA.

Alternatively, if AMY8e were included in the inversion, then it need not be duplicated. Instead, we would need to seek further repeats beyond this probe, and before the X-Y homologous regions, which are not involved in the inversion. It will be noted from the contig map that coverage of the region distal to AMY8e is only in CEPH YACs, with no linkage between the AMY8e and X-Y homologous regions in

OXEN YACs. Thus, it is possible that the inversion breakpoint falls in this region, although the lack of AMY8e sequences in any YACs from the alphoid contig is still problematic.

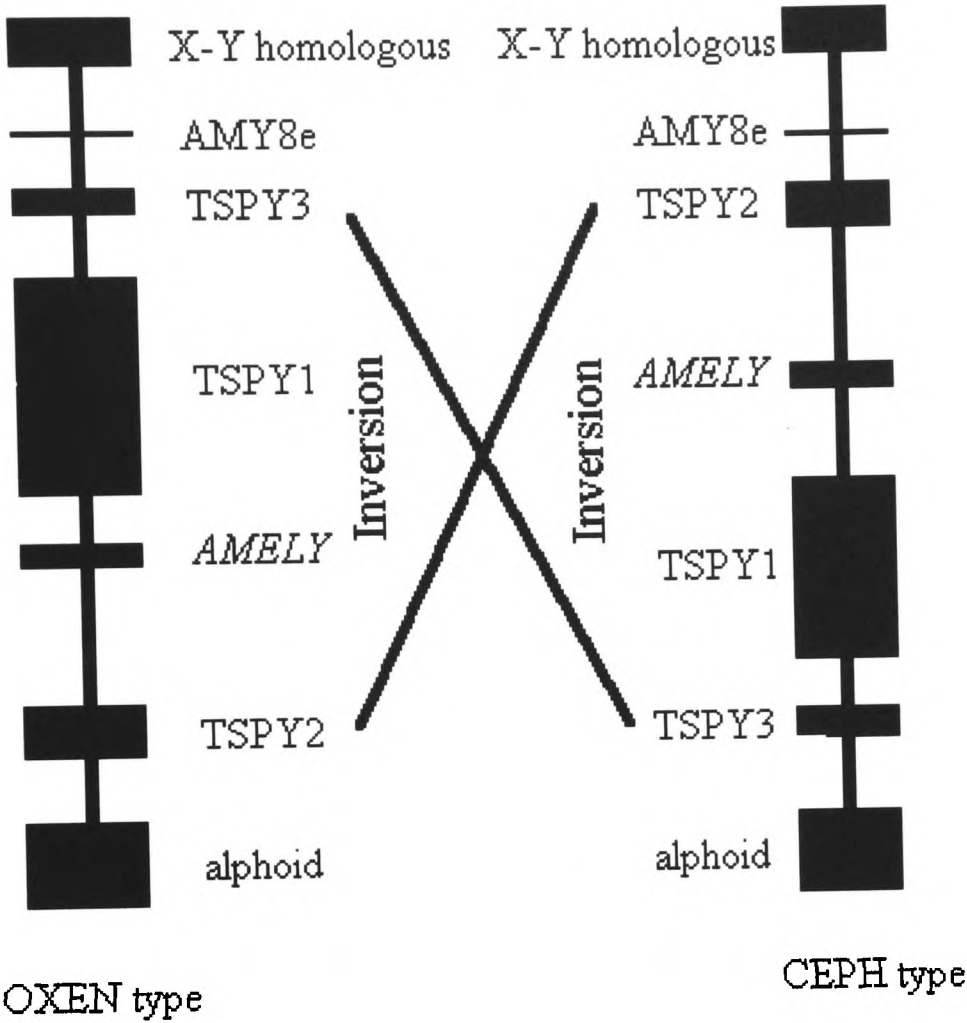


Figure 7.1: An illustration of the sequences on the short arm of the Y chromosome hypothesised to be involved in the inversion polymorphism.

An additional complexity is raised by the nature of the TSPY2 and 3 blocks – unpublished data (C. Tyler-Smith, pers. comm.) shows that these vary in size between the chromosome types – in OXEN chromosomes, the proximal block is 55 kb, while the distal block is 30 kb; CEPH chromosomes have the block sizes as 33 kb and 50 kb, respectively. These differing sizes suggest that the inversion of the TSPY blocks is not a simple recombination outwith the TSPY blocks themselves, else the blocks would be of identical sizes. Of course, it may be that each block, like the main TSPY1 block, is polymorphic in size between individuals, and this discrepancy merely reflects this polymorphism.

A third possibility for the inversion mechanism would be that the duplicated regions are not involved in the inversion at all, and the inversion occurs between these regions, only involving the *AMELY* and TSPY1 blocks.

The precise nature of this inversion is thus more complex than was originally thought, and at present we do not fully understand the nature of the polymorphism.

The breakpoints of the inverted region have not yet been accurately defined, although the duplicated region described above still seems to be the most likely candidate sequence. Experiments could be carried out to determine where the breakpoints lie, such as more detailed mapping of YAC or other clones from each of the types of chromosome from each of the putative breakpoints, to search for consistent discrepancies between the two types of chromosome.

Once the breakpoints are known in more detail, it may be possible to detect diagnostic differences between the two forms of the inversion by means of differences in restriction enzyme fragments from the breakpoints – a similar diagnostic test for the forms of the emerin inversion was used by Small & Warren. Such a test would be far simpler to use than the stretched FISH technique which was originally used to identify the inversion, and would enable a large number of different populations to be scored for the inversion.

One possible consequence of the inversion in evolutionary terms is discussed in the following section, in the context of sex-reversed individuals.

The third set of experiments concerned with the sequence features of the Y chromosome carried out in this work was an attempt to locate a number of reported genes on the YAC map. Several PCR assays were developed based on the published cDNA sequences of a number of novel genes. Screening of panels of YACs from the short arm or the long arm allowed several of these genes to be located on the map. On the short arm, the *PRKY* gene was localised to the *AMELY* contig, in the region involved in the inversion, and the *TTY1* and *TTY2* genes were co-localised with the TSPY-containing YACs from all three TSPY blocks. A cosmid containing only a TSPY repeat unit was also tested with the *TTY1* and *TTY2* primers, and confirmed that

the *TTY* sequences were not the same as the TSPY sequences, merely closely linked. A fourth putative short arm gene, *PRY*, was tested on the YAC panel, but was found not to be present on the region covered by the present map; this gene was also proposed as being present on the distal long arm (Lahn & Page 1997), and so the primers designed to detect it may have been specific for the long arm copy of the gene. Alternatively, the gene may well be present on the short arm, as suggested, but falls in the gap between the *AMELY* and the TSPY contigs. This does not seem unlikely, as Lahn & Page suggest that the short arm copy of *PRY* co-localises with the TSPY sequences.

Against this interpretation, that *PRY* 'slips through the cracks' as it were, stands the evidence given by a subsequent analysis of a pair of deleted Y chromosomes from normal males in Sri Lanka (described in Pandya 1998, and Santos, Pandya, & Tyler-Smith, 1998), which were noted as lacking the *AMELY* gene, although having no other obvious ill effects. Testing of these chromosomes with the *PRKY*, *PRY* and *TTY* genes demonstrated that, while *PRKY* was deleted, as expected from the proposed recombination event between the TSPY1 and TSPY2 blocks, *PRY* was still present in the deleted chromosomes. This suggests that the *PRY* gene does not lie between the two blocks of TSPY sequence; the evidence is, however, consistent with the suggestion that the *PRY* detected by these primers lies on the long arm only. The presence or absence of the two *TTY* genes was not mentioned in the previous sentences as this gave a complicated picture – *TTY2* was deleted from these chromosomes, while *TTY1* was still present. This suggests first of all that the two sequences are not as closely linked as was previously thought; and secondly that the actual organisation of these sequences with respect to the TSPY regions is complex. The localisation from the PCR tests carried out on the YAC panels was not very precise, showing only that the sequences were found on those YACs which also carried TSPY sequences. A more detailed mapping would be of great assistance in firmly placing these genes with respect to the three TSPY blocks.

In addition to these short arm genes, several genes were located to the long arm. *UTY*, *DBY* and *DFFRY* were all placed with the assistance of the YAC panels, *DFFRY* being the most proximal of the three. Two further genes were investigated, *TB4Y* and *BPY1*, but found to lie distal to the GMGY6 contig.

The location of an anti-Turner syndrome gene was undertaken with the assistance of a pair of ring Y chromosomes, DM and TS. Both chromosomes had all of the short arm sequences tested from *SRY* to *DYZ3*, while TS lacked all long arm loci tested, and DM had the *RBF5* long arm locus, but lacked *GMGXY6*, immediately distal to *RBF5*. Importantly, TS had been clinically determined as having Turner syndrome, while DM, although infertile (thanks to the loss of the majority of the long arm genes) was not a Turner patient (details of the cases are given in Tzancheva et al., in preparation). It was discovered that both individuals lacked the *DFFRY* gene, so excluding it as a potential Turner gene (as indeed did another study conclude [Brown et al., 1998], instead designating *DFFRY* as the *AZF_a* gene). This leads to the conclusion that there exist one or more genes on the long arm, proximal to the *DFFRY* locus, which are involved in protecting against Turner syndrome. These genes yet remain to be found and identified; obviously there is further work to be carried out here. Clusters of rare-cutting restriction enzyme sites in the YAC contig may mark the location of these genes.

7.3: The effect of sequence organisation on the occurrence of sex reversal.

The most common group of XX males have undergone an illegitimate recombination between the X and the Y chromosomes, with the breakpoint at the *PRKY* gene, which acts as a recombination hotspot between the chromosomes (Klink et al., 1995). The equivalent class of XY females, who have lost the short arm sequences distal to this point, is much less common, but still occurs. As the *PRKY* gene is included in the region defined by the inversion described above, it was thought that the form of the inversion may affect the production of such sex-reversed individuals. That is to say, since in order to have the recombination, both the *PRKX* and the *PRKY* genes must be in the same orientation, and as the inversion is not thought to occur on the X chromosome, only one form of the inversion will be able to give rise to this form of sex reversal. This is illustrated in figure 7.2 below.

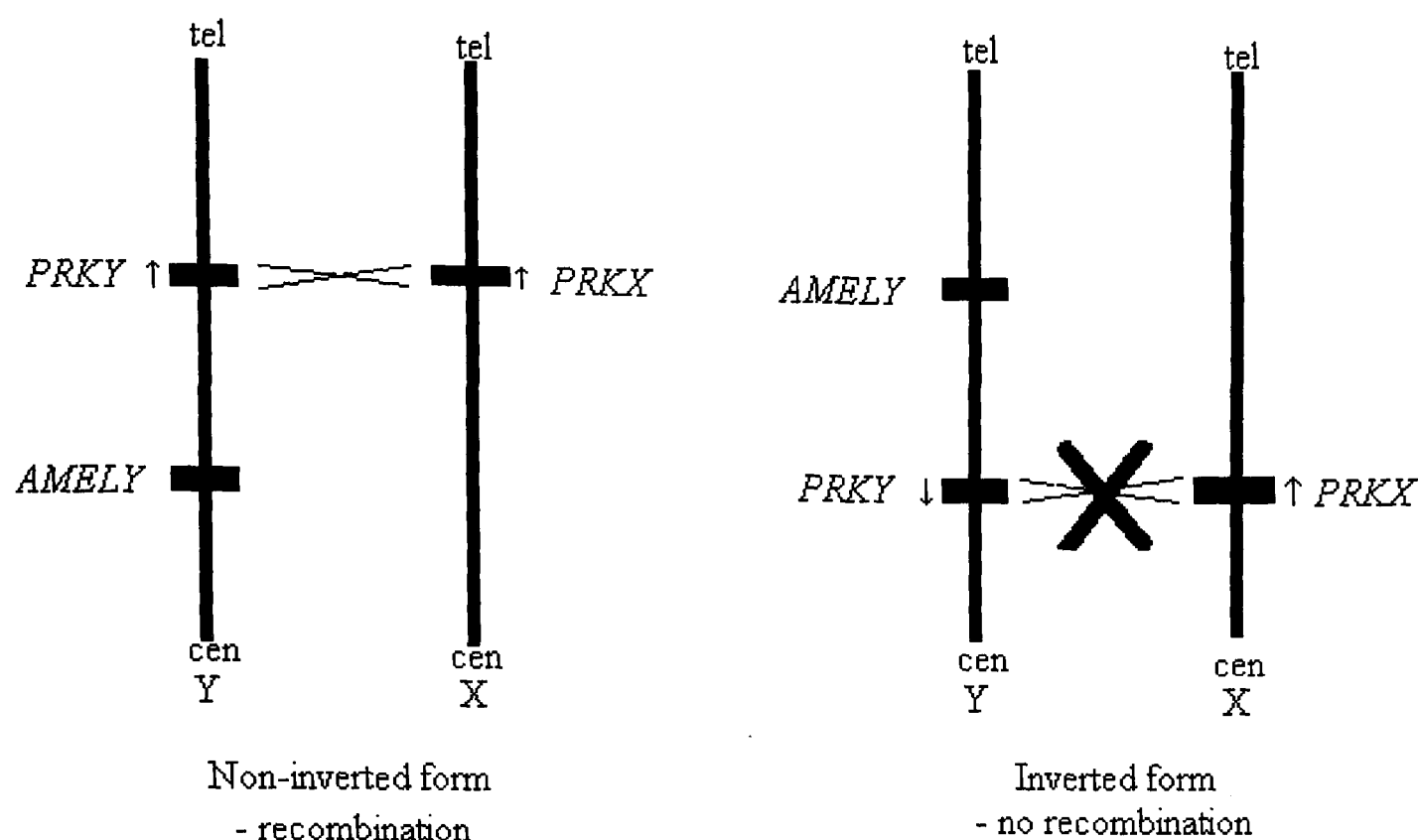


Figure 7.2: The hypothesised effect of the short arm inversion on occurrence of sex-reversing recombination between the sex chromosomes.

In order to test this hypothesis, a number of aberrant Y chromosomes of this type were tested for the Y chromosome polymorphisms defined by SRY-1532 (Kwok et al., 1996), 92r7, and YAP.

These polymorphisms defined the location of the Y chromosomes on the haplotype tree constructed by Pandya (1998), and shown in figure 6.15. The locations of the Boleth (CEPH) and OXEN chromosomes were known on this tree. Due to the nature of the chromosomal rearrangements, the other informative polymorphisms could be tested only in XY females or the fathers of XX males, who were seldom available.

However, the available data did show that all the XY females tested fell into 3 of the possible haplotypes, all of which were of the OXEN type (alphoid – *AMELY* – TSPY1); while the possible set of groups to which the XX males belonged was larger, but still a limited subset of the available haplotypes. The XX males could not be shown to belong to either the OXEN or the CEPH types alone, so it is still formally possible that the inversion is not restrictive for the generation of this class of XX males. However, given the information obtained from the XY females, and the (very small) sample of fathers of XX male individuals which were analysed – both of whom

fell into the same grouping as the XY females – it seems likely that all of these sex-reversed individuals with the *PRKY* breakpoint would fall into the same OXEN class of chromosomes. Obviously, further work on this problem is necessary; the limiting resource here being the scarcity of samples from the parents of the sex-reversed individuals.

Nonetheless, this result shows that a certain subset of Y chromosomes have an evolutionary advantage, in that it seems likely that the CEPH form of Y chromosome will not give rise to this particular class of sex-reversal. Of course, the CEPH chromosomes will still be affected by other types of sex-reversal, but as one form is prevented from arising on this background, the slight reproductive advantage gained by these individuals will be enough to allow selection to take place. The frequency of incidence of this class of XX male is 1 in 60,000 of the population (Schiebel et al., 1997). It will be recalled from chapter 5 that it was predicted that the OXEN form of chromosome was the most common; a bolder prediction can now be made that, barring random genetic drift or other advantages of the OXEN form, at some far future point the CEPH form of chromosome will come to predominate in the population.

The result obtained herein illustrates a concrete practical application of the mapping work carried out. In assembling the YAC map, the TSPY3 duplication was detected; having detected the duplication, a mechanism was available whereby the short arm inversion could occur; detection of the inversion made clear that the *PRKY* recombination hotspot was involved in the inversion; and this led to the conclusion that the orientation of the *PRKY* gene can directly affect the production of sex-reversed individuals, which in turn has effects on the evolution of the Y chromosome.

7.4 Implications of these results

7.4.1 The Human Genome Project

What are the implications of these results, with regard to the aims of this project, and the points raised in the introduction? First of all, take the Human Genome Project: while the goal of the HGMP is to map and eventually to sequence the entire human genome, a number of conclusions may be drawn from the data presented here. This

project has examined the organisation of a region of a single chromosome in more detail than would be usual for other parts of the HGMP, aside from the sequencing efforts; and it is unlikely that all of this region would be involved in any sequencing effort from this chromosome.

It was known that previous maps of this region were in error, due to the undetected presence of numerous repeated sequences; and so the current work was much more careful to avoid making errors of this type. However, it may be assumed that repetitive regions elsewhere in the genome would cause much the same problems as the Y chromosome repeats, and that these problems may be unnoticed unless the region is studied in more detail.

Most other regions of the genome under study are assumed to be unique regions, and so the problems of repeats would not arise. However, it should be noted that, as the Y chromosome was studied in more detail, repeats and duplications were found that were not suspected of being duplicated (i.e., the short arm duplication). If this result is typical for the genome, then it may be that other undetected duplications exist in the genome, and more detailed study would notice these regions. More importantly, if they remain undetected, then closure of maps will be assumed to exist where gaps remain. Thus, when such maps are used as the basis of further study, such as sequencing, large portions of the genome will not be present.

Finally, the short arm inversion detected here reflects a polymorphism between different Y chromosomes. How common are such polymorphisms on other chromosomes? And, if present, would they affect the HGMP? The Y chromosome polymorphism was not detected in the YAC libraries, and no discrepancies were observed between the libraries. However, if such inversions are present on other chromosomes, then they may be presumed to affect the mapping effort. Differences between clone libraries used to construct the maps would lead to multiple maps from the same region; meanwhile genetic maps would also show alternative results where different cell lines were used to construct the maps. Again, more serious is the case where such inversions are not detected, even as divergent maps – such would be the case for the Y chromosome, if FISH had not been used to confirm the inversion. In

such cases, data could be used which is not applicable to a subgroup of chromosomes, causing uninterpretable results to be obtained.

Obviously, the more closely a genomic region is studied, the more likely it is that features such as these duplications or inversions will be noticed, if they exist. Such features can seriously affect the understanding of the genome, and discredit the accuracy and consistency of previous maps of the region. The major question for the HGMP is, would detailed study of other regions of the genome also reveal such features as the Y chromosome does? This question cannot be answered here; the reliability of other chromosomal maps must await closer scrutiny. It may be considered that the Y chromosome is unique by virtue of its non-recombining portion, being more prone to accumulate duplications and inversions than other chromosomes. But it would be well to be aware of this problem when considering data from other chromosomes.

7.4.2 Centromeres and MACs

A second aspect of this work to consider is what does it tell us about centromeres? And, as a major goal of centromere studies is to gain information that may be of use in the construction of a MAC, does this project help reach that goal? Firstly, the centromere. Over the course of this work, the emphasis of centromere research has shifted from the idea of a particular 'centromere sequence' toward that of centromere complexes being maintained epigenetically, the sequences around the centromere are still important to understand. There may be some aspect of natural centromere DNA which causes centromeres to remain in that location, rather than drifting elsewhere; also natural centromeres are more stable than neocentromeres, so there may be some sequence aspect to the stability as well.

It will be recalled from section 1.7.2 that the short arm sequences adjacent to the alphoid DNA are considered important for full centromere function of the Y chromosome – some 300 kb of adjacent sequences, according to Tyler-Smith et al. (1993). This 300 kb includes the limits of the short arm inversion described in chapter 5. As there is no noticeable difference in segregation efficiency between the CEPH and OXEN types of chromosome, it must be concluded that any centromeric

sequences on the short arm lie closer to the alphoid DNA than the inversion breakpoint.

For construction of a MAC, it is essential to understand not only what sequences must be included to obtain a stable chromosome, but also how small such a MAC may be. The lower size limits of a MAC must be determined by how close to a functioning centromere genes may lie and still be expressed. From the current study, we have placed genes close to the centromere on both chromosome arms, so giving valuable information on this aspect of MAC construction.

On the long arm, the *DFFRY* gene lies approximately 3 Mb from the alphoid sequences (this cannot be precise thanks to the gaps which still exist in this part of the map). However, it is also known that there must be at least one gene active on the chromosome proximal to this, the anti-Turner syndrome gene. Thus, the closest long arm gene to the centromere is under 3 Mb away.

The short arm provides a more complex picture. Some *TTY* sequences can be found only 500 kb from the alphoid DNA, but it is not known if these are transcribed, there being several other *TTY* locations along the short arm. Indeed, it is not yet known if *TTY* is a functional gene, rather than just a transcribed stretch of DNA. Assuming that the closest *TTY* sequences are not functional genes, the closest functional region of the chromosome depends on which type it is. CEPH chromosomes have the functional TSPY1 region lying around 1 Mb from the alphoid DNA; while in OXEN chromosomes, the inversion places *AMELY* around 800 kb from the alphoid DNA. However, it may be significant that, while both *AMELY* and TSPY are transcribed, this occurs at only a low level; it may be that proximity to the centromere inhibits transcription. TSPY becomes even more interesting when its organisation is considered: TSPY is present in long tandem arrays, and is the only known human protein-coding gene to be organised in such large arrays. Why is this so? It could be a mechanism to compensate for transcriptional inhibition caused by its proximity to the centromere. *AMELY* would not need such a mechanism as it has an active X homologue, *AMELX*.

Thus, present knowledge places functional genes a minimum of 800 kb from the centromeric DNA; such a constraint is vital to know if a functional MAC is to be constructed. Of course, there may be no actual constraint to how close a gene may lie to the centromere, in which case further studies may be expected to reveal this. Telomere truncations within alphoid arrays have placed an expressed marker gene at the centromere (Brown et al., 1994), but the long-term stability of this expression is unknown.

7.5 Further work to be carried out

It will be noted that several points raised in this thesis could well be expanded upon were sufficient time and resources available. First of all, the final remaining gaps in the map could be closed. There is no reason to believe that closure would be impossible; a suitable clone library must be found from which to work, and diligent effort ought to produce mapping data to cover the gaps. As stated previously, though, this effort is unlikely to be brought to bear upon the gaps, unless some particularly pressing reason is found to devote the resources to the mapping effort. From the information obtained in this project, however, there do not seem to be many gaps of great interest – the *PRY* gene may lie in the short arm gap, although it may not be present on the short arm at all; and the majority of the long arm gaps are thought to be repetitive and satellite sequences.

The second avenue for exploration is the determination of the inversion breakpoints on the short arm. An effort to isolate clones from both chromosome sources covering the proposed breakpoints should give more information regarding sequence differences between the two types of chromosome.

Positive identification of the inversion breakpoints could then be used to develop a diagnostic test for the orientation of the inverted sequence (such as PCR or restriction fragment lengths, as used for the X chromosome emerin inversion [Small & Warren, 1998]). Such a test ought to be simpler and quicker to use than the FISH method used in this thesis, and could be applied to DNA other than tissue culture cell lines.

Such a test could, in turn, be applied to the study of the *PRKY* sex-reversed individuals. Perhaps all the affected individuals could be tested to determine the

orientation of the inversion directly, rather than relying on the limited availability of such suitable polymorphisms as were used in this thesis. This approach would help to confirm or deny the hypothesis that this entire class of sex-reversal only occurs on a background of the OXEN type, and never on the CEPH type of chromosome.

Finally, further study of the genes so far identified to lie in this pericentromeric region would be of interest. Identification of functions may be possible from individuals which have deleted certain genes. One or more of the gene(s) responsible for protecting against Turner syndrome are now known to lie proximal to *DFFRY* on the long arm; as yet, however, no potential candidates are known to be found on this part of the chromosome. Clustering of restriction sites on the map may suggest possible regions in which to concentrate the search for these genes (at least two such clusters are apparent proximal to *DFFRY* from the map); if no genes are found from here, then this may provide a reason to begin trying to close the gap between the 68 bp satellite contig and the 14A3C contig.

So, the work of this thesis has been to extend the contig maps to the limits of YAC cloning, and two of the gaps between contigs have been closed. After which, a short arm duplication was identified, along with a related short arm inversion. A number of genes were localised on both the short and long arms from the map, one of which, *PRKY*, was involved in the inversion. This led to a study of individuals with Y-chromosomal breakpoints at the *PRKY* region, to determine whether or not the orientation of the *PRKY* gene affected the occurrence of breakpoints. Finally, a summary map marking the most important features noted in this project is given below, in figure 7.3.

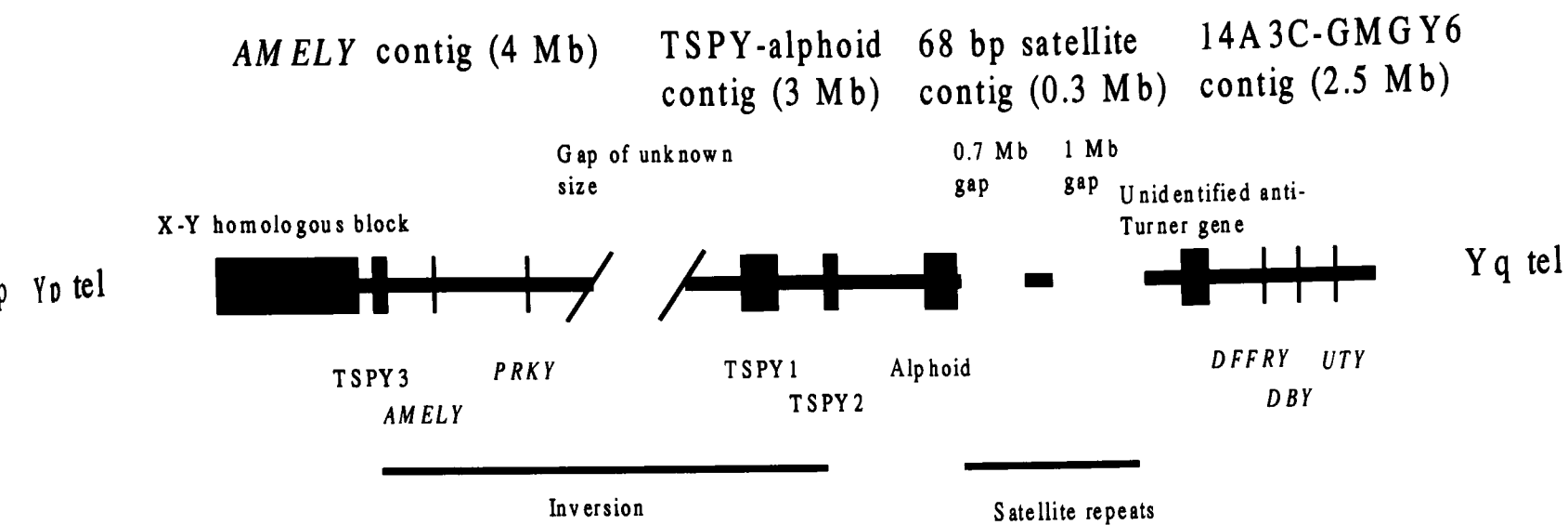


Figure 7.3: A final summary sketch map, showing the main features of the chromosome thus far mapped, and the remaining gaps in the contig map.

References

- Andersen LB, Tommerup N & Koch J (1990) Formation of a minichromosome by excision of the proximal region of 17q in a patient with von Recklinghausen neurofibromatosis. *Cytogenet. Cell Genet.* **53**: 206–210
- Affara NA, Floentin L Morrison N, Kwok K, Mitchell M, Cook A, Jamieson D, Glasgow L, Meredith L, Boyd E & Ferguson-Smith MA (1986a) Regional assignment of Y-linked DNA by deletion mapping and their homology with X-chromosome and autosomal sequences. *Nuc. Acids Res.* **14**: 5353-5373
- Affara NA, Ferguson-Smith MA, Tolmie J, Kwok K, Mitchell M, Jamieson D, Cook A & Florentin L (1986b) Variable transfer of Y-specific sequences in XX males. *Nuc. Acids Res.* **14**: 5375-5387
- Affara NA, Lau Y-FC, Briggs H, Davey P, Jones MH, Khwaja O, Mitchell M & Sargent C (1994) Report of the first international workshop on human Y chromosome mapping. *Cytogenet. Cell Genet.* **67**: 359-402
- Affara NA, Bishop C, Brown W, Cooke H, Davey P, Ellis N, Graves JM, Jones M, Mitchell M, Rappold G, Tyler-Smith C, Yen P & Lau Y-FC (1996) Report of the second international workshop on Y chromosome mapping. *Cytogenet. Cell Genet.* **73**: 33–76
- Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J, Kawasaki K, Minoshima S & Shimizu N (1997) Human BAC library: construction and rapid screening. *Gene* **191**: 69–79
- Barboux S, Vilain E, Raoul O, Gilgenkrantz S, Jeandidier E, Chadenas D, Souleyreau N, Fellous M & McElreavey K (1995) Proximal deletions of the long arm of the Y-chromosome suggest a critical region associated with a specific subset of characteristic Turner stigmata. *Hum. Mol. Genet.* **4**: 1565–1568

Bellane-Chantelot C, Lacroix B, Ougen P, Billault A, Beauvils S, Bertrand S, Georges I, Glibert F, Gros I, Lucotte G, Susini L, Codani J-J, Gesnouin P, Pook S, Vaysseix G, Lu-Kuo J, Ried T, Ward D, Chumakov I, Le Paslier D, Barillot E & Cohen D (1992) Mapping the whole human genome by fingerprinting YACs. *Cell* **70**: 1059-1068

Bickmore WA & Cooke HJ (1987) Evolution of homologous sequences on the human X and Y chromosomes, outside of the meiotic pairing segment. *Nuc. Acids Res.* **15**: 6261-6271

Birnboim HC & Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids Res.* **7**: 1513-1523

Bischoff FR, Maier G, Tilz G & Ponstingl H (1990) A 47 kDa human nuclear protein recognised by antikinetochore autoimmune sera is homologous with the protein encoded by RCC1. *PNAS* **87**: 8617-8621

Botstein D, White RL, Skolnik M & Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314-331

Brown GM, Furlong RA, Sargent CA, Erickson RP, Longepied G, Mitchell M, Jones MH, Hargreave TB, Cooke HJ & Affara NA (1998) Characterisation of the coding sequence and fine mapping of the human *DFFRY* gene and comparative expression analysis and mapping to the *Sxr^b* interval of the mouse Y chromosome of the *Dffry* gene. *Hum. Mol. Genet.* **7**: 97-107

Brown KE, Barrett MA, Burgtorf C, Shaw P, Buckle VJ & Brown WRA (1994) Dissecting the centromere of the human Y chromosome with cloned telomeric DNA. *Hum. Mol. Genet.* **3**: 1227-1237

Burke DT, Carle GF & Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**: 806-812

Cai M & Davis RW (1990) Yeast centromere binding protein CBF1, of the helix-loop-helix family, is required for chromosome stability and methionine prototrophy. *Cell* **61**: 437–446

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

Clarke L (1990) Centromeres of budding and fission yeast. *TIG* **6**: 150–154

Cohen D, Chumakov I & Weissenbach J (1993) A first generation physical map of the human genome. *Nature* **366**: 698–701

Connelly C & Hieter P (1996) Budding yeast *SKP1* encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. *Cell* **86**: 275–285

Cooke CA, Heck MS & Earnshaw WC (1987) The INCENP antigens : movement from inner centromere to midbody during mitosis. *J. Cell Biol.* **105**: 2053–2067

Cooke H (1976) Repeated sequence specific to human males. *Nature* **262**: 182–186

Cooper KF (1992) Analysis of the centromeric region of the human Y chromosome. D.Phil thesis, Oxford University

Cooper KF, Fisher RB & Tyler-Smith C (1992) Structure of the pericentromeric long arm of the human Y chromosome. *J. Mol. Biol.* **228**: 421–432

Cooper KF, Fisher RB & Tyler-Smith C (1993) Structure of sequences adjacent to centromeric alpha satellite on the human Y chromosome. *J. Mol. Biol.* **230**: 787–799

Demidov V, Frank-Kamenetskii MD, Egholm M, Buchardt O & Nielsen PE (1993) Sequence selective DNA cleavage by PNA targeting using nuclease S1. *Nuc. Acids Res.* **21**: 2103–2107

Depinet TW, Zackowski JL, Earnshaw WC, Kaffe S, Sekhon GS, Stallard R, Sullivan BA, Vance GH, van Dyke DL, Willard HF, Zinn AB & Schwartz S (1997) Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. *Hum. Mol. Genet.* **6**: 1195–1204

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES & Botstein D (1987) A genetic linkage map of the human genome. *Cell* **51**: 319–337

Dower WJ, Miller JF & Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nuc. Acids Res.* **16**: 6127–6145

du Sart D, Cancilla MR, Earle E, Mao J-I, Saffery R, Tainton KM, Kalitsis P, Martyn J, Barry AE & Andy Choo KH (1997) A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nature Genetics* **16**: 144–153

Ekwall K, Olsson T, Turner BM, Cranston G & Allshire RC (1997) Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. *Cell* **91**: 1021–1032

Ellison JW, Wardak Z, Young M, Robey P, Webster M & Chiong W (1997) PHOG, a candidate gene for involvement in the short stature of Turner syndrome. *Hum. Mol. Genet.* **6**: 1341–1347

Ferrin LJ & Camerini-Otero RD (1991) Selective cleavage of human DNA: RecA assisted restriction endonuclease (RARE) cleavage. *Science* **254**: 1494–1497

Ferrin LJ & Camerini-Otero RD (1994) Long range mapping of gaps and telomeres by RARE cleavage. *Nature Genetics* **6**: 379–383

Fisher EM, Beer-Romero P, Brown LG, Ridley A, McNeil JA, Lawrence JB, Willard HF, Bieber FR & Page DC (1990) Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* **63**: 1205-1218

Fisher RB (1992) Molecular studies of the human Y chromosome. D.Phil thesis, Oxford University

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

Freije D, Helms C, Watson MS & Donis-Keller H (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258**: 1784-1787

Frommer M, Prosser J & Vincent P (1984) Human satellite I sequences include a male-specific 2.47 kb tandemly repeated unit containing 1 Alu family member per repeat. *Nuc. Acids Res.* **12**: 2887-2900

Fukagawa T & Brown WRA (1997) Efficient conditional mutation of the vertebrate CENP-C gene. *Hum. Mol. Genet.* **6**: 2301-2308

Gaff C, du Sait D, Kalitsis P, Ianello R, Nagy A & Choo KHA (1994) A novel nuclear protein binds centromeric alpha satellite DNA. *Hum. Mol. Genet.* **3**: 711-716

Gietz D, St Jean A, Woods RA & Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. *Nuc. Acids Res.* **20**: 1425

Gläser B, Hierl T, Taylor K, Schiebel K, Zeitler S, Papadopoulos K, Rappold G & Schempp W (1997) High-resolution fluorescence *in situ* hybridisation of human Y-linked genes on released chromatin. *Chrom. Res.* **5**: 23-30

Guellaen G, Casanova M, Bishop C, Geldwerth D, Andre G, Fellous M & Weissenbach J (1984) Human XX males with Y single-copy DNA fragments. *Nature* **307**: 172-173

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J & Goodfellow PN (1996) A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**: 339–346

Haaf T, Warburton PE & Willard HF (1992) Integration of human alpha satellite DNA into simian chromosomes. *Cell* **70**: 681–692

Haaf T & Ward DC (1994a) High resolution ordering of YAC contigs using extended chromatin and chromosomes. *Hum. Mol. Genet.* **3**: 629–633

Haaf T & Ward DC (1994b) Structural analysis of alpha satellite DNA using extended chromatin and chromosomes. *Hum. Mol. Genet.* **3**: 697–709

Haldi M, Perrot V, Saumier M, Desai T, Cohen D, Cherif D, Ward D & Lander ES (1994) Large human YACs constructed in a rad52 strain show a reduced rate of chimaerism. *Genomics* **24**: 478–484

Halverson D, Baum M, Stryker J, Carbon J & Clarke L (1997) A centromere DNA-binding protein from fission yeast affects chromosome segregation and has homology to human CENP-B. *J. Cell Biol.* **136**: 487–500

Hammer MF & Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**: 951–962

Harrington JJ, van Bokkelen G, Mays RW, Gustashaw K & Willard HF (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nature Genetics* **15**: 345–355

Heiskanen M, Karlin R, Hellsten E, Peltonen L, Kallioniemi OP & Palotie A (1994) High resolution mapping using FISH to extended DNA fibres prepared from agarose embedded cells. *BioTechniques* **17**: 928–933

Heller R, Brown KE, Burgtorf C & Brown WRA (1996) Mini-chromosomes derived from the human Y chromosome by telomere directed chromosome breakage. *PNAS* **93**: 7125–7130

Heng HQH, Squire J & Tsui L-C (1992) High resolution mapping of mammalian genes by in situ hybridisation to free chromatin. *PNAS* **89**: 9509–9513

Henke A, Fischer C & Rappold G (1993) Genetic map of the human pseudoautosomal region reveals a high rate of recombination in female meiosis at the Xp telomere. *Genomics* **18**: 478–485

Hirst MC, Rack K, Nakahori Y, Roche A, Bell MV, Flynn G, Christadoulou Z, MacKinnon RN, Francis M, Littler AJ, Anand R, Poustka A-M, Lehrach H, Schlessinger D, D'Urso M, Buckle VJ & Davies KE (1991) A YAC contig across the fragile X site defines the region of fragility. *Nuc. Acids Res.* **19**: 3283-3288

Jeffreys AJ, Wilson V & Thein SL (1985) Hypervariable minisatellite regions in human DNA. *Nature* **314**: 67–73

Jiang W, Middleton K, Yoon H-J, Fouquet C & Carbon J (1993) An essential yeast protein, Cbf5p, binds in vitro to centromeres and microtubules. *Mol. Cell Biol.* **13**: 4884–4893

Jobling MA, Fretwell N, Dover GA & Jeffreys AJ (1994) Digital coding of human Y chromosomes: MVR-PCR at Y-specific minisatellites. *Cytogenet. Cell Genet.* **67**: 390

Jobling MA & Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *TIG* **11**: 449-456

Jones MH, Khwaja OSA, Briggs H, Lambson B, Davey PM, Chalmers J, Zhou C-Y, Walker EM, Zhang Y, Todd C, Ferguson-Smith MA & Affara NA (1994) A set of 97 overlapping YAC clones spanning the human Y chromosome euchromatin. *Genomics* **24**: 266–275

Kere J, Nagaraja R, Mumm S, Cicciodicola A, D'Urso M & Schlessinger D (1992) Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. *Genomics* **14**: 241–248

Kim UJ, Birren W, Slepak T, Mancino V, Boysen C, Kang HL, Simon MI & Shizuya H (1996) Construction and characterisation of a human bacterial artificial chromosome library. *Genomics* **34**: 213–218

Kipling D & Warburton PE (1997) Centromeres, CENP-B, and *Tigger* too. *TIG* **13**: 141–145

Kirsch S, Keil R, Edelmann A, Henegariu O, Hirschmann P, LePaslier D & Vogt PH (1996) Molecular analysis of the genomic structure of the human Y chromosome in the euchromatic part of its long arm (Yq11). *Cytogenet.. Cell Genet.* **75**: 197–206

Klink A, Schiebel K, Winkelmann M, Rao E, Horsthmeke B, Ludecke H-J, Claussen U, Scherer G & Rappold G (1995) The human protein kinase gene PKX1 on Xp22.3 displays Xp/Yp homology and is a site of chromosomal instability. *Hum. Mol. Genet.* **4**: 869–878

Kwok C, Tyler-Smith C, Mendonca BB, Hughes I, Berkovitz GD, Goodfellow PN & Hawkins JR (1996) Mutation analysis of the 2 kb 5' to SRY in XY females and XY intersex subjects. *J. Med. Genet.* **33**: 465–468

Lahn BT & Page DC (1997) Functional coherence of the human Y chromosome. *Science* **278**: 675–680

Lambson B, Affara NA, Mitchell M & Ferguson-Smith MA (1992) Evolution of DNA sequence homologies between the sex chromosomes in primate species. *Genomics* **14**: 1032–1040

Larin Z, Fricker MD & Tyler-Smith C (1994) De novo formation of several features of a centromere following introduction of a Y alphoid YAC into mammalian cells. *Hum. Mol. Genet.* **3**: 689–695

Lau EC, Mohandas TK, Shapiro LJ, Slavkin HC & Snead ML (1989) Human and mouse amelogenin loci are on the sex chromosomes. *Genomics* **4**: 162-168

Le MH, Duricka D & Karpen GH (1995) Islands of complex DNA are widespread in *Drosophila* centric heterochromatin *Genetics* **141**: 283-303

Le Y & Dobson MJ (1997) Stabilization of yeast artificial chromosome clones in a *rad54-3* recombination-deficient host strain. *Nuc. Acids Res.* **25**: 1248-1253

Lechner J & Carbon J (1991) A 240 kDa multisubunit protein complex, CBF3, is a major component of the budding yeast centromere *Cell* **64**: 717-725

Ma K, Inglis JD, Sharkey A, Bickmore WA, Hill RE, Prosser EJ, Speed RM, Thomson EJ, Jobling M, Taylor K, Wolfe J, Cooke HJ, Hargreave TB & Chandley AC (1993) A Y chromosome gene family with RNA-binding protein homology: Candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell* **75**: 1287-1295

Maniatis T, Fritsch EF & Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbour Press, New York

Masumoto H, Masukata H, Muro Y, Nozaki N & Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA. *J. Cell Biol.* **109**: 1963-1973

Meluh PB & Koshland D (1995) Evidence that the MIF2 gene of *S. cerevisiae* encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Mol. Biol. Cell* **6**: 793-807

Murphy DT & Karpen GH (1995) Localization of centromere function in a *Drosophila* minichromosome. *Cell* **82**: 599-609

Nakahori Y, Mitani K, Yamada M & Nakagome Y (1986) A human Y-chromosome specific repeated DNA family consists of a tandem array of pentanucleotides. *Nuc. Acids Res.* **14**: 7569-7580

Neil DL, Villasante A, Fisher RB, Vetrie D, Cox B & Tyler-Smith C (1990) Structural instability of human tandemly repeated DNA sequences cloned in YAC vectors. *Nuc. Acids Res.* **18**: 1421-1428

Nelson DL, Ballabio A, Victoria MF, Pieretti M, Bies RD, Gibbs RA, Maley JA, Chinault AC, Webster TD & Caskey CT (1991) Alu-primed PCR for regional assignment of 110 YACs from the human X chromosome. *PNAS* **88**: 6157-6161

Olson M, Hood L, Cantor CR & Botstein D (1989) A common language for physical mapping of the human genome. *Science* **245**: 1434-1435

Page DC, 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

Palazzolo MJ, Sawyer SA, Martin CH, Smoller DA & Hartl DL (1991) Optimised strategies for STS selection in genome mapping. *PNAS* **88**: 8034-8038

Palmer DK, O'Day K, Wener MH, Andrews BS & Margolis RL (1987) A 17 kDa centromere protein (CENP-A) copurifies with nucleosome core particles and histones. *J. Cell Biol.* **104**: 808-815

Pandya, A (1998) D.Phil thesis, Oxford University.

Parra I & Windle B (1993) High resolution visual mapping of stretched DNA by fluorescent in situ hybridisation. *Nature Genetics* **5**: 17-21

Pavan W, Hieter P, Sears D, Burkhoff A & Reeves RH (1991) High-efficiency yeast artificial chromosome fragmentation vectors. *Gene* **106**: 125-127

Rao E, Weiss B, Fukami M, Rump A, Niesler B, Mertz A, Muroya K, Binder G, Kirsch S, Winkelmann M, Nordsiek G, Heinrich U, Breuning MH, Ranke MB, Rosenthal A, Ogata T & Rappold GA (1997) Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature Genetics* **16**: 54–63

Rappold GA (1993) The pseudoautosomal regions of the human sex chromosomes. *Hum. Genetics* **92**: 315–324

Rattner JB (1991) The structure of the mammalian centromere. *BioEssays* **13**: 51–56

Reijo R, Lee T-Y, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O, de la Chapelle A, Silber S & Page DC (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding gene. *Nature Genetics* **10**: 383–393

Ried K, Mertz A, Nagaraja R, Trusgnich M, Riley JH, Anand R, Lehrach H, Page D, Ellison JW & Rappold G (1995) Characterization of a YAC contig spanning the pseudoautosomal region. *Genomics* **29**: 787–792

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

Santos FR & Tyler-Smith C (1995) PCR haplotypes for the human Y chromosome based on alphoid satellite variants and heteroduplex analysis. *Gene* **165**: 191–198

Santos FR, Pandya A & Tyler-Smith C (1998) Reliability of DNA-based sex tests. *Nature Genetics* **18**: 1

Sargent CA, Briggs H, Chalmers IJ, Lambson B, Walker E & Affara NA (1996) The sequence organisation of the Yp/proximal Xq homologous regions of the human sex chromosomes is highly conserved. *Genomics* **32**: 200–209

Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, Rozen S, Dinulos MB, Disteche CM & Page DC (1996) The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nature Genetics* **14**: 292–299

Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Disteche C, McGillivray B, de la Chapelle A & Page DC (1998) Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum. Mol. Genet.* **7**: 1–11

Small K & Warren ST (1998) Emerin deletions occurring on both Xq28 inversion backgrounds. *Hum. Mol. Genet.* **7**: 135–139

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503–517

Southern EM, Anand R, Brown WRA & Fletcher DS (1987) A model for the separation of large DNA fragments by crossed field gel electrophoresis. *Nuc. Acids Res.* **15**: 5925–5943

Steiner N & Clarke L (1994) A novel epigenetic effect can alter centromere function in fission yeast. *Cell* **79**: 865–879

Stoler S, Keith KC, Curwick KE & Fitzgerald-Hayes M (1995) A mutation in CSE4, an essential gene encoding a novel chromatin associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev* **9**: 573–589

Strobel SA & Dervan PB (1990) Site specific cleavage of a yeast chromosome by oligonucleotide directed triple helix formation. *Science* **249**: 73–75

Sugimoto K, Yata H, Muro Y & Himeno M (1994) CENP-C is a DNA binding protein which possesses a novel DNA binding motif. *J. Biochem.* **116**: 877–881

Sun X, Wahlstrom J & Karpen G (1997) Molecular structure of a functional *Drosophila* centromere. *Cell* **91**: 1007–1019

Tomkiel JA, Cooke CA, Saitoh H, Bernat RL & Earnshaw WC (1994) CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase. *J. Cell. Biol.* **125**: 531–545

Tyler-Smith C & Brown WRA (1987) Structure of the major block of alpha satellite DNA on the human Y chromosome. *J. Mol. Biol.* **195**: 457–470

Tyler-Smith C, Taylor L & Müller U (1988) Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *J. Mol. Biol.* **203**: 837–848

Tyler-Smith C, Oakey RJ, Larin Z, Fisher RB, Crocker M, Affara NA, Ferguson-Smith MA, Muenke M, Zuffardi O & Jobling MA (1993) Localisation of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nature Genetics* **5**: 368–375

Tyler-Smith C & Willard HF (1993) Mammalian chromosome structure. *Curr. Op. Gen. Dev.* **3**: 390–397

Tzancheva M, Kaneva R, Kumanov P, Tyler-Smith C & Williams G (1998) Two male patients with ring Y: Implications for refining the regions responsible for Turner syndrome. In preparation.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL & Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**: 996–1005

Vafa O & Sullivan KF (1997) Chromatin containing CENP-A and α -satellite DNA is a major component of the inner kinetochore plate. *Curr. Biol.* **7**: 897–900

Vergnaud G, Page DC, Simmler MC, Brown L, Rouyer F, Noel B, Botstein D, Chapelle A & Weissenbach J (1986) A deletion map of the human Y chromosome based on DNA hybridisation. *Am. J. Hum. Genet.* **44**: 756-765

Vogt PH, Edelmann A, Kirsch S, Heneguriu O, Hirschmann P, Kieseewetter F, Kohn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Grone H-J, Jung A, Engel W & Haidl G (1996) Human Y chromosome azoospermia factor (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet.* **5**: 933-943

Vogt PH, Affara N, Davey P, Hammer M, Jobling M, Lau Y-FC, Mitchell M, Schempp W, Tyler-Smith C, Williams G, Yen P & Rappold GA (1998) Report of the third international workshop on Y chromosome mapping. *Cytogenet. Cell Genet.* in press.

Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS & Page DC (1992) The human Y chromosome: A 43-interval map based on naturally occurring deletions. *Science* **258**: 52-59

Voullaire LE, Slater HR, Petrovic V & Choo KHA (1993) A functional marker centromere with no detectable alpha satellite, satellite 3 or CENP-B protein. *Am. J. Hum. Genet.* **52**: 1153-1163

Warburton PE, Cooke CA, Bourassa S, Vafa O, Sullivan BA, Stetten G, Gimelli G, Warburton D, Tyler-Smith C, Sullivan KF, Poirier GG & Earnshaw WC (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr. Biol.* **7**: 901-904

Weber JL (1990) Human DNA polymorphisms and methods of analysis. *Curr. Op. Biotech.* **1**: 166-171

Weil D, Wang I, Dietrich A, Poustka A, Weissenbach J & Petit C (1994) Highly homologous loci on the X and Y chromosomes are hot-spots for ectopic recombinations leading to XX maleness. *Nature Genetics* **7**: 414-419

Weller PA, Chritcher R, Goodfellow PN, German J & Ellis NA (1995) The human Y-chromosome homolog of XG – Transcription of a naturally truncated gene. *Hum. Mol. Genet.* **4**: 859–868

Whitfield LS, Hawkins TL, Goodfellow PN & Sulston J (1995) 41 kilobases of analyzed sequence from the pseudoautosomal and sex-determining regions of the short arm of the human Y chromosome. *Genomics* **27**: 306–311

Wiegant J, Kalle W, Mullenders L, Brookes S, Hoovers JM, Dauwerse JG, van Ommen GJ & Raap AK (1992) High resolution in situ hybridisation using DNA halo preparations. *Hum. Mol. Genet.* **1**: 587–591

Yen TY, Li G, Schaar BT, Szilak I & Cleveland DW (1992) CENP-E is a putative kinetochore motor that accumulates just before mitosis. *Nature* **359**: 536–539

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjiddmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgratov OV & Tyler-Smith C (1997) Genetic relationship of Asians and Northern Europeans revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**: 1174–1183

Zinn AR, Page DC & Fisher EMC (1993) Turner syndrome: the case of the missing sex chromosome. *TIG* **9**: 90–93

Abbreviations

AMELY – amelogenin, Y copy

ARS – autonomously replicating sequence

ATP – adenosine triphosphate

AZF – azoospermia factor

BAC – bacterial artificial chromosome

bp – base pair

BPY1 – basic protein Y

BSA – bovine serum albumen

CBF – centromere binding factor

CDE – centromeric DNA element

CENP – centromere protein

CEPH – centre d'étude polymorphisme humain

cM – centimorgans

CREST – calcinosis, Raynaud's phenomenon, oesophageal dysmotility, sclerodactyly, telangiectasia

CTP – cytosine triphosphate

DAPI - diamidino-2phenylindole dihydrochloride

DAZ – deleted in azoospermia

DBY – DEAD box Y

DFFRY – *Drosophila* fat-facets related, Y copy

DMSO – dimethyl sulphoxide

DNA – deoxyribonucleic acid

DTT - dithiothreitol

EDTA – ethylenediaminetetraacetic acid

FISH – fluorescent *in situ* hybridisation

FITC – fluorescein isothiocyanate

GBY – gonadoblastoma on the Y

GCY – growth control on the Y

GTP – guanidine triphosphate

HGMP – human genome mapping project

HGMP-RC – HGMP resource centre
INCENP – inner CENP
kb – kilobases
LDS – lithium dodecyl sulphate
MAC – mammalian artificial chromosome
Mb – megabases
mRNA – messenger RNA
NTP – nucleotide triphosphate
PABY – pseudoautosomal boundary on the Y
PAC – P1-derived artificial chromosome
PAR – pseudoautosomal region
PBS – phosphate buffered saline
PCR – polymerase chain reaction
PEG – polyethylene glycol
PFGE – pulsed-field gel electrophoresis
PRKY – protein kinase, Y copy
RFLP – restriction fragment length polymorphism
RNA – ribonucleic acid
SDS – sodium dodecyl sulphate
SRY – sex-determining region of the Y
STS – sequence tagged site
TE – Tris-EDTA buffer
TSPY – testis-specific protein on the Y
TTP – thymidine triphosphate
TTY – testis transcript Y
UTY – ubiquitous TPR motif Y
YAC – yeast artificial chromosome
YAP – Y Alu polymorphism
ZFY – zinc finger Y