Human Molecular Genetics, 2000, Vol. 9, No. 17 **2563–2572**

Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events

Christine Kamp, Peter Hirschmann, Hartmut Voss[1], Karin Huellen and Peter H. Vogt[+]

Reproduction Genetics, Institute of Human Genetics, University of Heidelberg, Im Neuenheimer Feld 328 and [1]LION-Bioscience, D-69120 Heidelberg, Germany

Received 3 August 2000; Revised and Accepted 25 August 2000

DDBJ/EMBL/GenBank accession nos AF290421–AF290423

We mapped the breakpoints of the AZoospermia factor a (AZFa) microdeletion located in proximal Yq11 in six men with complete germ cell aplasia, i.e. Sertoli Cell Only syndrome (SCO). The proximal breakpoints were identified in a long retroviral sequence block (HERV15yq1: 9747 nucleotides) at the 5′ end of the DYS11 DNA locus in Yq11, interval D3. The distal breakpoints were found in a homologous HERV15 sequence block mapped to the Yq11 interval D6, i.e. in the distal part of the AZFa region (HERV15yq2: 9969 nucleotides). Compared with the HERV15yq1 sequence, HERV15yq2 is marked by a deletion of a HERV15 sequence domain at its 5′ end and insertion of an LINE 1 3′-UTR sequence block (L1PA4) of similar length at its 3′ end. The deletion of the L1PA4 element was recognized as the molecular origin of the DYS11 12f2 restriction fragment length polymorphism. For all six AZFa patients it was possible to perform PCR experiments bridging both retroviral sequence blocks, which map in a distance of 781.557 kb in proximal Yq11 in fertile men. The AZFa breakpoint–fusion regions were located in their recombined HERV15yq1/HERV15yq2 sequence blocks in either one of two long identical sequence domains (ID1 and ID2). We therefore assume that intrachromosomal recombination events between the two homologous retroviral sequence blocks in proximal Yq11 are probably the causative agents for most of the AZFa microdeletions observed in men with SCO syndrome. A mean value of 792 kb was estimated for their molecular lengths.

INTRODUCTION

DNA deletions are common mutational events. Their size can vary from a single base to several megabases of DNA. Small DNA deletions (<20 bp) are mediated by short repeats (1). Their deletion frequency is proportional to the size of the repeat and to the extent of their homology, and inversely proportional to the molecular distance between them. Deletions of DNA regions resulting in the loss of megabases of DNA occur with a high frequency in specific genomic regions commonly characterized by homologous locus-specific clusters of repetitive sequence elements which include transcriptionally active pseudogenes (2). Microdeletions on autosomes can cause complex pathological phenotypes also described as contiguous gene syndromes (3).

Three different microdeletions on the human Y chromosome cause male infertility (4). These were mapped in the euchromatic part of the long Y arm and designated as 'AZoospermia Factors': AZFa, AZFb and AZFc (5) because men who have one of these microdeletions in the Y chromosome suffer from azoospermia, i.e. mature sperm cells (spermatozoa) are not detectable in their seminal fluid. Of the men seeking help at infertility clinics ~15% suffer from this severe pathological phenotype (6) and in 5–20% of these patients the molecular analysis revealed a *de novo* microdeletion of at least one of the three AZF regions in their Y chromosomes (4). Based on the analysis of >1000 men with idiopathic azoospermia, we estimated the occurrence of a *de novo* microdeletion in Yq11 in the German male population to be at a rate of ~10^{-4}. Even more Yq11 microdeletions may be present in some Italian male populations, because of those analysed >55% of men with Sertoli Cell Only (SCO) syndrome were diagnosed as having a microdeletion in AZFa (7).

The AZFa region in proximal Yq11 is defined by the deletion of the Y chromosomal sequence tagged site (STS) loci: sY83, sY84, sY85, sY86 and sY87 (5). This corresponds to the deletion of Yq11 intervals D3–D6 according to the map of Vogt *et al.* (5) and to the deletion of Yq11 interval 5C according to the map of Vollrath *et al.* (8), respectively. But naturally, as the number of STS loci mapped per Y deletion interval does not correlate to its molecular extension, STS deletion mapping is not suitable for the analysis of the molecular extension of the AZFa deletion in proximal Yq11. This first became possible after the alignment of a series of sequenced bacterial artificial chromosome (BAC) clones from the human Y chromosome forming a contig in the AZFa region (9). However, the causative agents which so frequently create these Y chromosomal microdeletions in the human population remained unknown.

By sequence analysis of the DYS11 locus in proximal Yq11 we identified a long retroviral sequence block of ~10 kb, which

[+]To whom correspondence should be addressed. Tel: +49 6221 565050; Fax: +49 6221 563710; Email: peter_vogt@med.uni-heidelberg.de

was strongly related to the human endogenous retrovirus 15 (HERV15). As this block was present in both proximal and distal AZFa breakpoint regions, we explored the possibility that it might be involved in the induction of AZFa deletions as a result of homologous intrachromosomal recombination events. For this purpose, we developed a series of new STS markers around both retroviral sequence blocks (HERV15yq1 and HERV15yq2) and with appropriate PCR experiments were able to bridge the breakpoint–fusion region of six AZFa patients. The breakpoint–fusion sites were mapped to either one of two long identical sequence domains (tentatively designated as 'ID1' and 'ID2') identified in the aligned HERV15yq sequence blocks. We therefore assume that these retroviral sequence blocks of the DYS11 locus are putative causative agents for the occurrence of the AZFa microdeletion in the majority of cases in this infertile patient group.

RESULTS

DYS11 locus is composed of a complex retroviral sequence structure

DYS11 was originally described as a DNA marker locus in interval 5 of the long arm of the human Y chromosome (10). Its marker probe, 12f3 (11), cross-hybridized to a male-specific genomic *Eco*RI fragment of 5.1 kb. Another marker probe, 12f2, proximal to 12f3 (11), cross-hybridized to polymorphic human Y-specific *Eco*RI fragments of 5.2 and 3.2 kb whose presence was mutually exclusive (12). Molecular deletion mapping with the 12f3 probe localized DYS11 to Yq11 interval D3 (5). This is in the proximal border region of AZFa microdeletions including deletion of Yq11 intervals D3–D6 of the Vogt map or interval 5C of the Vollrath map (13). AZFa microdeletions were found to be *de novo* mutation events in some men with a complete germ cell aplasia, Sertoli Cell Only (SCO) syndrome.

As the DYS11 probe 12f3 cross-hybridized to testis-specific transcripts in human and mouse (11), we sequenced its male-specific *Eco*RI fragment, p12f, in order to reveal the sequence of a putative AZFa candidate gene and deposited it in GenBank (accession no. AF290421). From sequence position 1 to 2384, we found a retroviral sequence structure with strong homology to the sequence of HERV15 (98% sequence identity), a human endogenous retrovirus (sequence published in RepBase: http://www.girinst.org), ending with a complete long terminal repeat (LTR) of this HERV sequence family (LTR15). Between positions 340 and 1254, a sequence block of *Homo sapiens* retinoic acid-inducible endogenous retroviral transcript (HSRIRT; GenBank accession no. M64936) was identified, pointing to a retroviral recombination event in this sequence region (Fig. 1a). HSRIRT is a retroviral transcript of an HERV sequence which belongs to the same sequence family as HERV15 (14). This family is characterized by a primer binding site for the 3′ end of isoleucine tRNA and therefore also called HERV-I family (15).

A copy of the autosomal highly conserved protein (*AHCP*) gene (GenBank accession no. AF097027) between two sequence blocks homologous to HERVK14, another human retrovirus (sequence published in RepBase: http://www.girinst.org), forms the second part of the 5107 nucleotide long DYS11 male-specific *Eco*RI fragment (Fig. 1a). The autosomal *AHCP* gene copy was mapped to the short arm of chromosome 6 (6p22–23) and encodes a putative protein of 413 amino acids (GenBank accession no. AF097027). The same *AHCP* coding frame was identified between sequence positions 2555 and 3765 in the DYS11 *Eco*RI fragment (Fig. 1a). However, some in-frame stop codons and frameshift mutations destroyed its coding potential for a Y chromosomal AHCP protein. The *AHCP* Y copy must therefore be classified as a pseudogene. Comparison

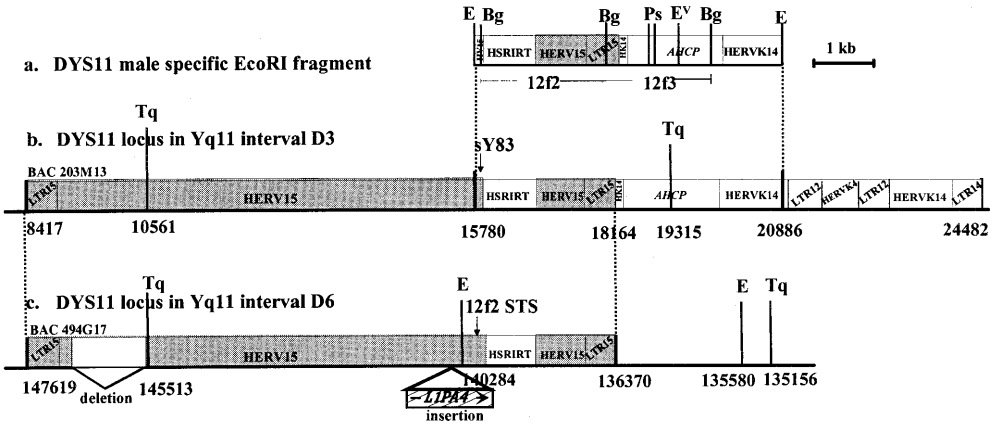

Figure 1. Schematic comparison of the retroviral sequence structures in DYS11 male-specific *Eco*RI fragment (**a**), in the DYS11 locus in the Yq11 interval D3 (**b**) and in the DYS11 locus in the Yq11 interval D6 (**c**). The DYS11 locus in the Yq11 interval D6 is specified by deletion of a HERV15 segment at its 5′ end and insertion of a LINE1 element (L1PA4) at its 3′ end in opposite orientation (c). In (b) and (c), sequence positions in the corresponding BAC clones are given below the block structures. The numbers are reversed for BAC 494G17 because the Y insert in this BAC clone has a reverse orientation (when compared with the orientation in Yq11). Positions of the two DYS11 STS markers are given above the blocks in (b) and (c). In (a) are given the restriction sites for *Eco*RI (E), *Bgl*II (Bg), *Pst*I (Ps) and *Eco*RV (E^V) and the extensions of the *Bgl*II fragments 12f2 and 12f3 (11). In (b) and (c) the sequence positions of the *Eco*RI (E) and *Taq*I (Tq) restriction sites are given which mark the polymorphic *Eco*RI and *Taq*I fragments cross-hybridizing to the 12f2 probe.

of the *Bgl*II, *Eco*RV and *Pst*I restriction map of the sequenced DYS11 *Eco*RI fragment with the published restriction data of Leroy *et al.* (11) confirmed the genuine structure of the sequence analyzed.

To find out whether the HERV15 and HERVK14 structures of the DYS11 locus extend beyond its male-specific *Eco*RI fragment, we aligned this sequence to the BAC sequence contig of the AZFa region published recently (9) and identified its position at the 5′ end of BAC 203M13 (GenBank accession no. AC002992) between nucleotides 15780 and 20886 (Fig. 1b). Alignment of the complete HERV15 sequence (8035 nucleotides) to BAC 203M13 revealed an extension of their homology in the DYS11 locus along the whole sequence length with a complete LTR15 sequence at positions 8417–8899 representing its 5′ end (Fig. 1b).

Distal to the DYS11 *Eco*RI fragment, the alignment of a complete HERVK14 sequence (5945 nucleotides) to the BAC 203M13 sequence extended the 3′ end homologous DYS11–HERVK14 region and ended with a complete LTR of this retrovirus (LTR14A) with 95% sequence identity in sequence positions 24140–24482 (Fig. 1b). The HERVK14 sequence homology was interrupted between positions 20985 and 22971 (2874 and 4758 of the HERVK14 sequence) due to insertion of a truncated HERVK4 homologous retroviral element in reverse orientation flanked by two LTR12 sequence blocks. Obviously, the DYS11 locus is composed mainly of a complex retroviral sequence structure.

HERV15 sequence structure of the DYS11 locus is duplicated in distal AZFa region

Two STS markers of the DYS11 locus are known: sY83 (8) and 12f2 (16). However, we only found sY83 in the BAC 203M13 sequence (positions 15836–16112), i.e. in the proximal AZFa region, as expected. We therefore searched for the position of the 12f2 STS locus in the complete BAC contig of the AZFa region (9). To exclude possible BAC sequence gaps or rearrangements we repeated this search also in our YAC contig (17) and in our panel of infertile men with various terminal deletions in Yq11 (5). STS 12f2 was mapped to BAC 494G17, to YAC 711F4 and to Yq11 interval D6 of our patient panel, i.e. always to the distal border region of the AZFa microdeletion including Yq11 intervals D3–D6 (5,17). As STS sY83 was mapped to Yq11 interval D3, i.e. to the proximal AZFa border region, two different locations of the DYS11 locus in proximal Yq11, near the proximal and distal AZFa breakpoint regions, were suggested.

We further explored this assumption by aligning the DYS11 locus sequence from BAC 203M13 (positions 8417–24482) to the complete AZFa BAC contig and identified its HERV15 homologous sequence domain in BAC 494G17 between sequence positions 147619 and 136370 (Fig. 1c: sequence numbers are reversed because the Y-sequence of this BAC had a reverse sequence orientation in the AZFa contig). The HERVK14 homologous sequence domain and the *AHCP* gene copy present at the 3′ end of the proximal DYS11 locus (Fig. 1b) could not be found in BAC 494G17 nor in any other BAC sequence of the AZFa region. The deletion of a short HERV15 sequence block in the 5′ end was also different (HERV15 sequence positions 228–1505) as well as the insertion of a truncated LINE 1 sequence designated as L1PA4 (18) in opposite orientation in the 3′ end (BAC 494G17 sequence positions 138212–139730) (Fig. 1c). The 12f2 STS locus was identified in the sequence region 138199–138626, i.e. with its forward primer in the unique L1PA4 sequence domain. This explains its presence only in the distal DYS11 locus (Fig. 1c). STS locus sY83 was not found in the BAC 494G17 sequence. Obviously, the two DYS11 STS markers are useful to mark specifically the proximal (sY83) and distal (12f2) DYS11 loci in proximal Yq11. We deposited the sequence of both DYS11-related HERV15 sequence blocks, tentatively designated as HERV15yq1 (9747 nucleotides from Yq11 interval D3) and HERV15yq2 (9969 nucleotides from Yq11 interval D6), in GenBank (HERV15yq1 and HERV15yq2, accession nos AF290422 and AF290423, respectively).

Internal HERV15yq structures and molecular origin of the 12f2 Y-DNA polymorphism

HERV15yq1 and HERV15yq2 are both complete human endogenous retroviral DNA sequences (HERVs). HERV genomic structures are basically analogous to those of the exogenous retroviruses. They contain coding information for: (i) *gag*, the group specific antigens, which are the capsid protein genes; (ii) *pol*, the polymerase gene encoding a protease, reverse transcriptase, RNase H and integrase protein; and (iii) *env*, the envelope protein gene. However, the HERVs' coding frames are usually defective because of numerous frameshift mutations and in-frame stop codons (15). The primer binding site for a specific tRNA located adjacent to the 5′ LTR is used as a classification marker for the multiple HERV sequence families. Together with a typical polypurine sequence tract (PPT) just in front of the 3′ LTR, both DNA sites serve as 'plus' and 'minus' templates for the retroviral DNA synthesis (15).

Adjacent to the putative 5′ LTRs of HERV15yq1 and HERV15yq2 there is a primer binding site complementary to the 3′ end of isoleucine tRNA and a PPT block of eight nucleotides was found at the beginning of their putative 3′ LTR blocks (Fig. 2). Both retroviral sequences therefore belong to the HERV-I subfamily (15). Alignment of the HERV15yq1 and HERV15yq2 sequence blocks and extensive BLAST sequence homology searches revealed that the typical retroviral coding blocks were also present in both HERV15yq sequences although their exact borderlines could not be estimated because of extensive divergences. Their detailed analysis was therefore not carried out. The putative 5′ end LTR15 sequence blocks of HERV15yq1 and HERV15yq2 displayed a 91% sequence identity, their putative 3′ end LTR15 sequence blocks were almost completely identical (99.7%) with only three single nucleotide mutations (SNMs) (Fig. 2). After the 5′ LTR their sequence identity decreased to 82% until it reached the HERV15 deletion site of the distal HERV15yq2 block (Fig. 1c). This HERV15 sequence part (HERV15 sequence 228–1505) is unique for the proximal HERV15yq1 block and was looped out (subsequently referred to as 'loop 1' with 1328 nucleotides) for further optimal alignment with the HERV15yq2 sequence (Fig. 2). Interestingly, loop 1 is followed by an identical sequence region of 1278 nucleotides ('ID1') after which numerous single nucleotide mutations again distinguish both HERV15yq sequence blocks

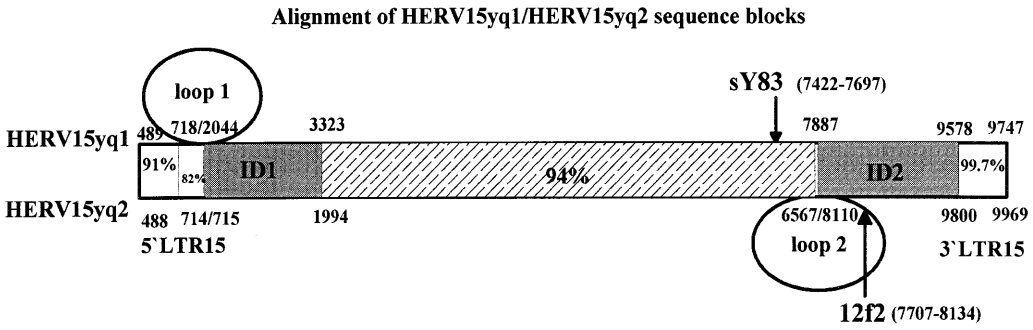

Figure 2. Schematic drawing of the alignment of the HERV15yq1 and HERV15yq2 sequences points to the presence of two identical sequence domains (ID1 with 1278 nucleotides and ID2 with 1690 nucleotides) after looping out the HERV15 segment deleted in HERV15yq2 (loop1 with 1328 nucleotides) and the L1PA4 element inserted in HERV15yq2 (loop 2 with 1543 nucleotides). HERV15yq1 sequence positions are given above the aligned block, HERV15yq2 sequence positions are given below the block. Sequence positions of the DYS11 marker sY83, only present in HERV15yq1, and marker 12f2, only present in HERV15yq2, are indicated. The percentage of sequence identities between HERV15yq1 and HERV15yq2 outside the ID segments are indicated separately for the 5′ LTR and 3′ LTR and the intervening sequence region in the middle of the block.

(beginning with position 3324 of HERV15yq1, corresponding to position 1995 of HERV15yq2).

The primer sequences for STS sY83 were only present in the proximal HERV15yq1 block (positions 7422–7697). The HERV15yq2 sequence then displayed a unique insertion of a truncated LINE 1 sequence, L1PA4, 6567–8110. L1PA4 was identified according to the classification of Smit *et al.* (18) by its specific nucleotide substitution pattern. It contained the 5′ forward primer of the 12f2 STS at positions 7707–7732 (Fig. 2). For further optimal alignment with the HERV15yq1 sequence the L1PA4 sequence part was also looped out ('loop 2' with a length of 1543 nucleotides). After loop 2 both HERV15yq sequence blocks displayed a second long identical sequence domain ('ID2' with 1690 nucleotides) which was only interrupted by the three single nucleotide mutations, after position 9578 in HERV15yq1 (9800 in HERV15yq2), in the 3′ LTR (Fig. 2).

Based on these alignment data, the molecular origin of the 12f2 STS polymorphism is assumed to be linked to the deletion of the L1PA4 element in the HERV15yq2 sequence block. The polymorphic 12f2 Y-DNA marker was first described as a restriction fragment length polymorphism (RFLP) for *Eco*RI and *Taq*I in human male DNA (12). In genomic DNA blots of human males and females, the probe 12f2 (Fig. 1a) cross-hybridized to a polymorphic male-specific *Eco*RI fragment of 5.2 and 3.2 kb, whose presence was mutually exclusive. Similarly, cross-hybridization to a polymorphic 10 kb *Taq*I fragment was observed. Its absence was related to duplication of the cross-hybridizing 8 kb *Taq*I fragment (12). A restriction map for *Eco*RI and *Taq*I around the HERV15yq2 sequence in BAC 494G17 revealed the presence of a 4.704 kb *Eco*RI fragment and 10.357 kb *Taq*I fragment as the putative cross-hybridizing fragments with the 12f2 probe, because both fragments include the 12f2 sequence (Fig. 1c). A restriction map for *Eco*RI and *Taq*I around the HERV15yq1 sequence in BAC 203M13 pointed to a 5.107 kb *Eco*RI fragment (the male-specific DYS11 *Eco*RI fragment) (Fig. 1a) and an 8.754 kb *Taq*I fragment (Fig. 1b) as putative 12f2 cross-hybridizing fragments.

We estimated a length of 1543 nucleotides for the inserted L1PA4 element in HERV15yq2. Its deletion would reduce the length of the 12f2 cross-hybridizing 10.357 kb *Taq*I fragment to 8.814 kb and of the 4.704 kb *Eco*RI fragment to 3.161 kb. These length reductions correspond well to those of the polymorphic 12f2 *Eco*RI and *Taq*I fragments described by Casanova *et al.* (12). We therefore assume that the molecular origin of the polymorphic 12f2 hybridization pattern first observed 15 years ago by Casanova *et al.* (12) is the deletion of the L1PA4 element in HERV15yq2. This view is also supported by the recently observed Y haplogroup 9 distinguished by a deletion of the 12f2 STS locus from Y haplogroup 2 (13).

Intrachromosomal HERV15yq1/HERV15yq2 recombinations cause AZFa microdeletions

We do not know of any function of HERV15yq1 and HERV15yq2 in the male germline. However, the two identical sequence domains (ID1 and ID2) of >1 kb in length (Fig. 2) suggest that they might be able to align intrachromosomally after meiotic breakage events in the proximal Yq11 region for subsequent local repair (19), thereby eventually deleting the intervening Y DNA region by some kind of recombination mechanism.

By using the genomic DNA samples of six patients (ANDSCH, ERWTHO, GERKEL, HERREI, JOLAR and LECGER) diagnosed as having a *de novo* AZFa deletion covering Yq11 intervals D3–D6 according to Vogt *et al.* (5), we explored this possibility experimentally by performing their precise AZFa breakpoint mapping. For this purpose we first developed a series of new STS markers proximal and distal to each HERV15yq sequence block: STSs AZFa-prox1 and AZFa-prox2 in BAC 203M13, STSs AZFa-dist1 and AZFa-dist2 in BAC 494G17 (Table 1). We were of the opinion that, if the Y chromosomal breakage–fusion event which results in the deletion of the AZFa region in these six patients took place in or near the HERV15yq1/HERV15yq2 sequence blocks, then the STSs AZFa-prox1 and AZFa-dist2 should always be present and the STSs AZFa-prox2 and AZFa-dist1 always absent. Our mapping results are presented in Figure 3a. We found that indeed all AZFa patients were positive for the STSs AZFa-prox1 and AZFa-dist2 and negative for the STSs AZFa-prox2 and AZFa-dist1. We therefore could conclude

Table 1. STS markers for AZFa breakpoint mapping

STS	Forward primer (5′→3′)	Reverse primer (5′→3′)	STS position in BAC	PCR product length	Annealing temperature (°C)
AZFa-prox1	AZFa-prox1-for: cttaaatgttgactcttcacc	AZFa-prox1-rev: gccttgtgtagaataagcagtca	203M13: 8176–8301	126 bp	58
sY83	sY83L: cttgaatcaaagaaggccc	sY83R: caatttggtttggctgacat	203M13: 15 836–16 112	275 bp	62
AZFa-prox2	AZFa-prox2-for: ggttcctgaacaggggact	AZFa-prox2-rev: ggcagcagaagggcctctc	203M13: 18 216–18 435	220 bp	68
AZFa-dist1	AZFa-dist1-for: ggcttctagtagtatggtc	AZFa-dist1-rev: ttgtccttcaatgcagatg	494G17: 148 020–147 631	390 bp	63
12f2	12f2-for: ctgactgatcaaaatgcttacagatc	12f2-rev: tcttctagaatttcttcacagaattg	494G17: 138 199–138 626	428 bp	59
AZFa-dist2	AZFa-dist2-for: gttccccattcattatactgttagc	AZFa-dist2-rev: gcactccagaaagataatacatc	494G17: 136 224–135 954	271 bp	65
AZFa-ID1[a]	203BP-7 (A): agctgaaactacttcttcagtttc	494BP-4 (B): gcacctcctttcccccagtaatg	(A): 203M13: 10 296–10 319 (B): 494G17: 144 235–144 257	1.55 kb[a]	61
AZFa-ID2[b]	sY83L (A): cttgaatcaaagaaggccc	AZFa-dist2-rev (B): gcactccagaaagataatacatc	(A): 203M13: 15 836–15 854 (B): 494G17: 135 954–135 976	2.8 kb[b]	61

[a]STS and PCR products found only in AZFa patients with breakpoint–fusion site in ID1 (Fig. 3b).
[b]STS and PCR products found only in AZFa patients with breakpoint–fusion site in ID2 (Fig. 3b).

that the Y chromosomal break and fusion points of the six individuals analysed must be located in or near the homologous HERV15yq sequence blocks.

Two patients (HERREI and JOLAR) were analysed positive for the proximal DYS11 STS sY83 but negative for the distal DYS11 STS 12f2, the other four (ANDSCH, ERWTHO, GERKEL and LECGER) were negative for sY83 but positive for 12f2 (Fig. 3a). The exact positions of the break–fusion sites in the HERV15yq1/HERV15yq2 sequence blocks of HERREI and JOLAR were therefore analysed by bridging PCR experiments with the forward primer of sY83 and the reverse primer of the AZFa-dist2 STS (Fig. 3b). A bridging PCR product of 2.8 kb was obtained from both genomic DNA samples. This amplification product was not present in the normal male and female control samples (Fig. 3b). We cloned both PCR products (HER-BP, JOL-BP) for sequence analysis and mapped the break and fusion regions for both AZFa patients to the ID2 domain of their recombined HERV15yq1/HERV15yq2 sequence blocks. Loop 2, specific for the HERV15yq2 block (Fig. 2), was not found in the fused HERV15yq1/HERV15yq2 sequence blocks of HERREI and JOLAR, but there was a continuous HERV15yq1 sequence identified by its diagnostic SNM pattern up to the start of the identical ID2 region. Interestingly, no sequence alteration could be found in the fused ID2 sequence along its complete length (1690 nucleotides). This suggests a precise intrachromosomal recombination mechanism between the HERV15yq1 and HERV15yq2 sequence blocks in the Y chromosome of both AZFa patients. Consequently, we were not able to identify the exact location of the break–fusion events in the ID2 sequence domains of HERREI and JOLAR. But since we found all three SNMs diagnostic for the 3′ LTR of the HERV15yq2 sequence distal of ID2, their recombination sites could only be located in the ID2 sequence domain.

Based on these results, we speculated whether similar recombination events are located near or in ID1 in the Y chromosome of the AZFa patients, ANDSCH, ERWTHO, GERKEL and LECGER. ID1 is the first long identical sequence domain of both aligned HERV15yq sequence blocks distal to loop 1 (Fig. 2). As the Y chromosomes of this second AZFa patient group still contained the distal 12f2 STS locus indicative for presence of the loop 2 structure in their fused HERV15yq1/HERV15yq2 sequence blocks, a recombination site different from that in HERREI and JOLAR was suggested (Fig. 3a). We therefore tested the possibility of a break–fusion event in the ID1 sequence domain for these patient samples by another breakpoint bridging PCR experiment. For a forward primer we used a sequence stretch of loop 1 unique for the HERV15yq1 sequence and directly flanking the ID1 domain (203BP-7), for a reverse primer, a sequence stretch specific for HERV15yq2 directly located distal to ID1 (494BP-4) was used. Genomic amplification products of the expected lengths (1.55 kb) were indeed found for all four patient samples (Fig. 3b). No PCR products were present in the male and female genomic control samples. We therefore assume for this second AZFa patient group (absence of STS sY83, presence of STS 12f2) also a recombination event of the HERV15yq1 and HERV15yq2 blocks as causative agent of their AZFa deletions, the break and fusion point now mapped to the ID1 domain.

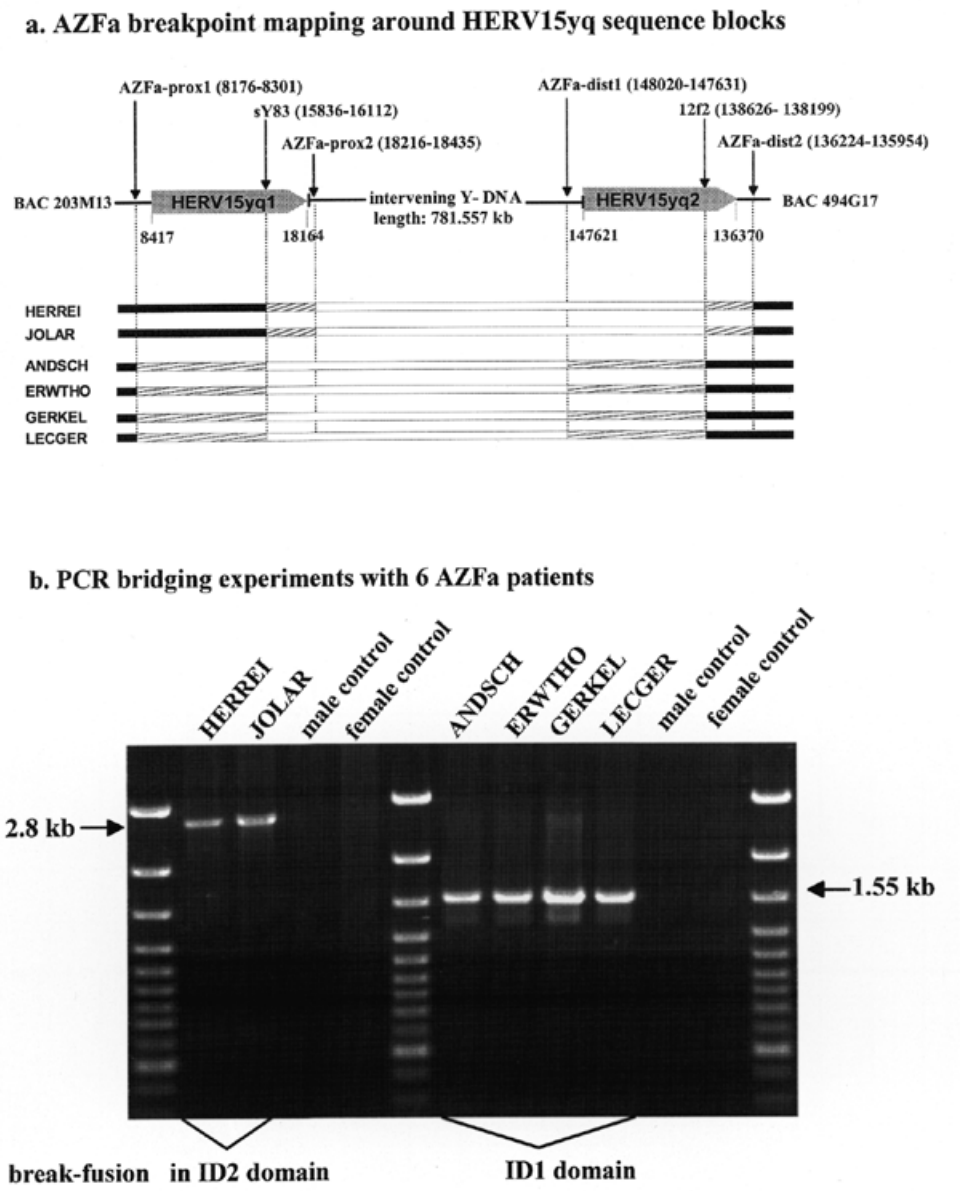

Figure 3. AZFa breakpoint mapping around HERV15yq1 and HERV15yq2 sequence blocks (**a**) inside them by PCR bridging experiments (**b**) in six AZFa patients. Two different break–fusion sites in the recombined HERV15yq1/HERV15yq2 blocks of the six AZFa patients were established. The Y chromosome of HERREI and JOLAR broke and fused in the ID2 domain, the Y chromosome of ANDSCH, ERWTHO, GERKEL and LECGER broke and fused in the ID1 domain. Primers used for bridging the AZFa breakpoints in ID1 were 203BP-7 and 494BP-4 (i.e. STS: AZFa-ID1), in ID2 sY83L and AZFa-dist2-rev (i.e. STS: AZFa-ID2) were used (Table 1).

DISCUSSION

Human endogenous retroviruses possibly originated from ancient germ-cell infections by exogenous retroviruses due to reverse transcription and genomic integration of the complementary DNA strands, or from ancestral retroviral elements due to subsequent transposition and recombination events (20). HERVs are stably integrated in the human genome and consequently inherited as Mendelian traits, although for several HERV sequences highly variable copy numbers (from 1 to >1000 per haploid genome) have been reported in various individuals (21). It is believed that most HERVs are transcriptionally silent and likely to be present in the human lineage since it diverged from the Old World monkeys, i.e. >25 million years ago. Interestingly, the Y chromosome seems to be the preferred location for different HERV sequence families (22), which might reflect the low density of protein coding genes on this male-specific chromosome and the absence of recombination along most of its length.

In this paper we describe the location and sequence structure of two homologous HERV15 sequence blocks in proximal Yq11 (tentatively designated as HERV15yq1 and HERV15yq2). Intrachromosomal recombination events between these loci

are probably the cause of the AZFa microdeletions in this Y region. Based on the AZFa BAC contig (9) a molecular distance of 781.557 kb between the two HERV15 sequences in proximal Yq11 could be estimated (Fig. 3a). This calculation became reasonable only after access to the extended N-terminal sequence of BAC 484O7 (GenBank accession no. AC006565; edition 20 July 2000), which now overlaps to the end of BAC 494G17 (positions 215726–223725 in GenBank accession no. AC005820) with a length of 8 kb (earlier only 1.5 kb) and therefore can be included in the AZFa BAC contig sequence (GenBank accession no. NT_001402). The molecular lengths of the AZFa deletions are dependent on their HERV15yq1/HERV15yq2 recombination sites. If these are in the ID1 region, we calculate for the AZFa deletion a size of 792.855 kb by addition of 781.557 kb plus 2044 bp of HERV15yq1 and 9969 bp minus 715 bp of HERV15yq2 (Fig. 2). If the recombination sites are in the ID2 region, we calculate a size of 791.304 kb by addition of 781.557 kb plus 9747 bp of the HERV15yq1 sequence. We therefore expect a length of ~792 kb as a mean value of an AZFa microdeletion in proximal Yq11.

HERV15yq1 and HERV15yq2 were identified as part of the polymorphic DYS11 locus in proximal Yq11 containing a complex retroviral DNA structure (Fig. 1). The sequence composition of the DYS11 DNA locus probably evolved from multiple retroviral insertion events, the order of which cannot be given. Most likely, the HERV15 sequence block was duplicated in proximal Yq11 before the pseudogene copy of the *AHCP* gene from chromosome 6 was transposed distal to HERV15yq1 in Yq11 interval D3. The distal end of HERV15yq2 coincided with the distal end of a X–Y homologous block of 60 kb in Yq11 interval D6 (data not shown). Further analysis will show whether the X chromosome contains also a duplication of this HERV15 sequence block.

Can DYS11 retroviral sequence structure promote genomic mutations in proximal Yq11?

We constructed the genomic sequence of the complete AZFa region by joining the overlapping sequence ends of the AZFa BAC contig (9) and analysed this Y region for similar complex retroviral sequence structures. Numerous LTRs of other HERV retroelements were identified as single or composite structures or flanking internal HERV sequences, but these retroelements were never as long as the described DYS11 retroviral sequence structure mapped in proximal Yq11, interval D3.

This raised the question as to whether the DYS11 retroviral sequence structure might be involved in the production of AZFa microdeletions in this Y region. Most AZFa deletions reported in the literature seem to include Yq11 intervals D3–D6 (4) although exceptional cases were reported as well (7,23,24). Prominent sequence elements, which might induce internal deletion or recombination events, are sequence stretches with homology to the chi recombination signal of *Escherichia coli* (GCTGGTGG), to immunoglobulin-like switch signals (TGGGG), or to the autonomous replication origins in yeast (ARS: A/TAAACATAAAA/T) (1). We found a GCTGGAGG stretch in the L1PA4 element of HERV15yq2 and numerous GCTGGGG stretches along the whole DYS11 DNA structure. ARS homology was found in the ID1 domain of HERV15yq1/HERV15yq2 at position 2204/874 with TAAACATAAAG. According to similar findings in the literature, these sequence elements suggest the possibility of internal recombinational hotspots in the DYS11 DNA structure. Longer sequence elements for internal recombination events are the repetitive LTR15 sequence blocks. We also identified a TATAA promoter box in the 5′ LTR of HERV15yq1 and HERV15yq2 and a canonical mRNA polyadenylation site (AATAAA) before the start of the 3′ LTR in both HERV15yq elements, suggesting transcriptional activity of the DYS11 locus. However, although similar transcription signals were observed in other HERVs, transcriptional activities of HERVs from the 5′ LTR promoter element were not yet observed in normal human tissues (R. Löwer, personal communication).

Which sequence features of the DYS11 structure, if any, are able to initiate the observed intrachromosomal recombination events of the HERV15yq1 and HERV15yq2 sequence blocks can only be found by further experiments. In this context, we note that most YACs which we tried to isolate from the AZFa region rearranged quickly during their preparative growth phases (17). Generally, one can assume that any molecular rearrangements in DYS11 leading to single strand or double strand breakage events will be able to initiate the observed recombinations between the HERV15yq1 and HERV15yq2 sequence blocks.

Large repetitive sequence blocks and genomic microdeletions

Homologous recombination events between large repetitive sequence blocks seem to be causative agents for the occurrence of genomic microdeletions in a number of different chromosomes (2). About 90% of patients with deficiency of the steroid sulfatase (STS) in Xp22.3 have a deletion of the entire *STS* gene due to intrachromosomal recombinations between copies of a repeat called S232 located 1.9 Mb apart in Xp22.3 (25). Interchromosomal homologous recombinations between a 24 kb repeat termed Charcot–Marie–Tooth disease type 1A repeat (CMT1A-REP) cause microdeletions of 1.5 Mb in 17p12 (26). Low copy repeats as putative agents for the induction of interstitial genomic microdeletions were also discussed for the Hunter syndrome (mucopolysaccharidosis type II) in Xq27.3–28 (27), for the velo-cardio-facial syndrome (VCFS) and DiGeorge syndrome (DGS) in 22q11 (28), for the Prader–Willi (PWS) and Angelman (AS) syndromes in 15q11–13 (29), for the Williams syndome (WS) in 7q11.23 (30) and for neurofibromatosis type 1 (NF1) in 17q11.2 (31). Most of these repeats contain a complex structure of tandem repetitive sequences with embedded transcriptionally active pseudogenes.

Sequence analyses of the breakpoint–fusion regions in the 24 kb CMT1A-REP in 17p12 map their recombination sites to long identical sequence stretches (>400 nucleotides) (32), resembling those we found at the recombination sites of the six AZFa breakpoint–fusion regions. The two putative recombination sites in the fused HERV15yq1/HERV15yq2 sequence structure of each of the six AZFa patients analysed were mapped in four AZFa patients to ID1, an identical sequence domain of 1278 nucleotides, and to ID2, an identical sequence domain of 1690 nucleotides, in two AZFa patients (Fig. 3). Sequence lengths of 132–232 nucleotides of perfect shared sequence identity were estimated as minimal efficient processing segments (MEPSs) for homologous recombinations

in mammalian meiosis (33). Long identical sequence stretches between recombined repeats seem therefore to be the general MEPS regions for inter- or intrachromosomal recombination events.

DYS11 12f2 polymorphism and AZFa deletions

The two DYS11 STS loci, sY83 (8) and 12f2 (16), were found to be markers for the proximal (sY83) and for the distal (12f2) HERV15yq sequence structure (Fig. 2). Deletion of the truncated LINE 1 element, L1PA4, in the HERV15yq2 sequence block was identified as the putative molecular origin of the 12f2 RFLP (12) and its polymorphic STS (16). It seems likely that the presence of the L1PA4 element in the HERV15yq2 sequence block represents the ancestral state in humans since populations with only the 'long' 12f2 alleles (5.1 kb *Eco*RI/ 10 kb *Taq*I) were found in African Blacks, in Orientals and in Native Americans and a frequency gradient of its deletion was observed in European populations (34). We assume that a gene conversion event between the two HERV15yq sequence blocks in proximal Yq11 most likely caused the precise deletion of this L1PA4 element.

According to the scheme of Y haplogroups published recently (13), men with a deletion of the 12f2 STS were classified as having a Y chromosome of haplogroup 9 (hg9). However, we found Y haplogroup 1 (hg1) for the Y chromosomes of JOLAR, ANDSCH, ERWTHO and LECGER and Y haplogroup 3 (hg3) for the Y chromosome of HERREI (16). This is consistent with the frequency of these haplogroups in unselected northern European populations (35), and the haplogroup defined by the 12f2 deletion (hg9) does not appear to be associated with the occurrence of an AZFa deletion. This view is supported by the occurrence of a high frequency of AZFa deletions in northern Italy (7,24) where hg9 is rare. However, most of these deletions were reported to be restricted to the Dead Box Y (*DBY*) gene (24) and are therefore certainly not caused by intrachromosomal HERV15yq1/HERV15yq2 recombination events. Different molecular mechanism(s) must induce these gene-specific deletion events, although it should be mentioned that the unusual high frequency of *DBY* deletions could not be confirmed in other Italian studies. Complete AZFa deletions were found with a lower frequency (36) more comparable to that found in other European countries (4).

MATERIALS AND METHODS

AZFa patients

The four AZFa patients, ANDSCH, GERKEL, ERWTHO and JOLAR have already been described (5,37). Similarily, patients HERREI and LECGER were first diagnosed in the routine AZF–PCR multiplex experiments (5) as having an AZFa deletion, i.e. a deletion of Yq11 intervals D3–D6. Their medical records indicated no reasons for their infertility and their karyotypes were proved to be normal (46,XY). Their genomic DNA samples were prepared by routine methods from fresh blood samples collected in EDTA tubes.

AZFa breakpoint mapping

STS primer pairs prepared for AZFa breakpoint mapping were designed from the sequences of BAC 203M13 (GenBank accession no. AC002992) and BAC 494G17 (GenBank accession no. AC005820)

Experimental conditions for their PCR amplification were optimized on the Biometra *T* Gradient Personal Cycler. Primer sequences and specific PCR annealing temperatures for each STS are listed in Table 1. All PCR experiments were carried out in a reaction volume of 50 µl of the PCR buffer of Gibco with 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 mM forward primer, 0.2 mM reverse primer; 2 U of Taq DNA polymerase (Gibco) and 200 ng of the genomic DNA samples.

PCR profile in the Biometra *T*-Gradient Personal Cycler

The conditions for 35 cycles of PCR were: pre-soak for 3 min at 94°C, denaturation for 1 min at 94°C, annealing for 1 min, at primer-specific annealing temperature (Table 1), polymerization for 1 min at 72°C and final extension for 2 min at 72°C.

The PCR products were analysed on a 1% agarose gel stained with ethidium bromide. As a length marker we used the GeneRuler 100 bp DNA ladder-plus and the GeneRuler 1 kb DNA ladder from MBI Fermentas.

PCR amplification and cloning of AZFa break–fusion regions of HERREI and JOLAR

To bridge the putative break–fusion regions of the AZFa patients HERREI and JOLAR with the primer pair sY83L and AZFa-dist2-rev [i.e. STS AZFa-ID2 (Table 1)], we improved our PCR buffer by increasing the dNTPs to 0.3 mM and the concentration of the *Taq* DNA polymerase to 3 U. The polymerization time was increased to 3 min.

The obtained PCR products of 2.8 kb were excised from a preparative 1% agarose gel and eluted with the Qiagen QIAquick Gel Extraction kit (Hilden, Germany). For the TA cloning experiment of the purified PCR fragments we used the pCR2.1-TOPO vector and the TOPO TA cloning kit from Invitrogen (Groningen, The Netherlands). Positive clones (HER-BP and JOL-BP) were identified by insert lengths and their DNA prepared from a 30 ml overnight luria broth culture using the NucleoBond PC100 kit (Macherey–Nagel, Dueren, Germany).

DNA sequence analysis

The male-specific DYS11 *Eco*RI fragment of the Y cosmid clone no. 12 (38) cloned in the *Eco*RI site of the pBR322 plasmid vector (p12f) was sequenced by using the Sanger dideoxy chain termination method and the strategy of primer walking on overlapping DNA subfragments. Both DNA strands were analysed. For sequence homology analyses in the database, restriction mapping and the analyses of putative protein coding frames we used the HUSAR software package of the German Cancer Research Center, Heidelberg (http://genome.dkfz-heidelberg.de/biounit).

The clones of the AZFa breakpoint–fusion plasmids of HERREI and JOLAR (HER-BP and JOL-BP) were sequenced on the Pharmacia A.L.F. express (Amersham Pharmacia Biotech Europe, Freiburg, Germany) using the Sanger dideoxy chain termination method [ALFexpress AutoRead Sequencing kit (Amersham Pharmacia Biotech)] and the strategy of primer walking on overlapping subfragments from both DNA strands. All sequence analyses were carried out at least twice from both

directions to exclude any sequence alterations in the fused HERV15yq1/HERV15yq2 sequence block only present in one sequence run. Database searches and analysis of the BAC sequences in the AZFa BAC contig (9) were performed with the BLAST programs of the National DC Center of Biotechnology Information, Washington, DC (http://www.ncbi.nlm.nih.gov). Multiple sequence alignments for mapping of the AZFa break and fusion region in the ID1/ID2 sequence domains of the fused HERV15yq1/HERV15yq2 blocks were carried out with the HUSAR software package (malign, clustal) of the German Cancer Research Center, Germany.

ACKNOWLEDGEMENTS

We are grateful to Mark Jobling for having provided us with the primer sequences of the 12f2 STS and its experimental PCR conditions. We thank Jean Weissenbach for gifting the p12f plasmid. We thank Ann Chandley and Tim Hargreave for the genomic DNA sample of JOLAR. Martin Hartmann is thanked for blood samples of the AZFa patients GERKEL, HERREI and ERWTHO, Frank Köhn for the blood sample of ANDSCH, Frank Kiesewetter for the blood sample of LECGER. We are indebted to Bart Janssen who carried out our routine AZF–PCR Multiplex program for LECGER. We also wish to thank Mrs Kartrin Christophers-Genzmer and Mrs Anne Jordan for their help in preparing the final version of this manuscript and especially for critically improving our English expression and grammar. This study was supported by a grant to P.H.V. from the Deutsche Forschungsgemeinschaft (DFG: Vo403/11.3).

REFERENCES

1. Krawczak, M. and Cooper, D.N. (1991) Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum. Genet.*, **86**, 425–441.
2. Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.*, **14**, 417–422.
3. Chen, K.-S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A.C., Lee, C.C. and Lupski, J.R. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for common contiguous gene deletion syndrome. *Nature Genet.*, **17**, 154–163.
4. Vogt, P.H. (1998) Human chromosome deletion in Yq11 AZF candidate genes and male infertility: history and update. *Mol. Hum. Reprod.*, **4**, 739–744.
5. Vogt, P.H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Köhn, F.M., Schill, W.B., Farah, S., Ramos, C. *et al.* (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet.*, **5**, 933–943.
6. Hargreave, T.B. (1990) Introduction. In Hargreave, T.B. and Soon, T.E. (eds), *The Management of Male Infertility*. P.G. Publishing, Singapore, pp. 2–21.
7. Foresta, C., Ferlin, A., Garolla, A., Moro, E., Postorello, M., Barbauy, S. and Rossato, M. (1998) High frequency of well-defined Y-chromosome deletions in idiopathic Sertoli cell-only syndrome. *Hum. Reprod.*, **13**, 302–307.
8. Vollrath, D., Foote, S., Hilton, A., Brown, L.G., Beer-Romero, P., Bogan, J.S. and Page, D.C. (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science*, **258**, 52–59.
9. Sun, C., Skaletzky, H., Birren, B., Devon, K., Tang, Z., Silber, S., Oates, R. and Page, D.C. (1999) An azoospermic man with a *de novo* point mutation in the Y chromosomal gene *USP9Y*. *Nature Genet.*, **23**, 429–432.
10. Vergnaud, G., Page, D.C., Simmler, M.-C., Brown, L., Rouyer, F., Noll, B., Botstein, D., de la Chapelle, A. and Weissenbach, J. (1986) A deletion map of the human Y chromosome based on DNA hybridization. *Am. J. Hum. Genet.*, **38**, 109–124.
11. Leroy, P., Seboun, E., Mattei, M.G., Fellous, M. and Bishop, C.E. (1987) Testis-specific transcripts detected by a human Y-DNA-derived probe. *Development*, **101**, 177–183.
12. Casanova, M., Leroy, P., Boucekkine, C., Fellous, M., Purrello, M., Fiori, G. and Siniscalco, M. (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science*, **230**, 1403–1406.
13. Vogt, P.H., Affara, N., Davey, P., Hammer, M., Jobling, M.A., Lau, Y.-F.C., Mitchell, M., Schempp, W., Tyler-Smith, C., Williams, G., *et al.* (1997) Report of the third international workshop on Y chromosome mapping 1997. *Cytogenet. Cell. Genet.*, **79**, 1–20.
14. Kannan, P., Buettner, R., Pratt, D.R. and Tainsky, M.A. (1991) Identification of a retinoic acid-inducible endogenous retroviral transcript in the human teratocarcinoma-derived cell line PA1. *J. Virol.*, **65**, 6343–6348.
15. Wilkinson, D.A., Mager, D.L. and Leong, J.-A. (1994) Endogenous human retroviruses. In Levy, J.A. (ed.), *The Retroviridae*, Vol. 3. Plenum Press, New York, NY, pp. 465–548.
16. Paracchini, S., Stuppia, L., Gatta, V., Palka, G., Moro, E., Foresta, C., Mengual, L., Ballesva, J.L., Oliva, R., Kremer, J.A.M. *et al.* (2000) Y-chromosomal DNA haplotypes in infertile European males carrying Y-microdeletions. *J. Endocrinol. Invest.*, **23**, in press.
17. Kirsch, S., Keil, R., Edelmann, A., Henegariu, O., Hirschmann, P., LePaslier, D. and Vogt, P.H. (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.
18. Smit, A.F.A., Tóth, G., Riggs, A.D. and Júrka, J. (1995) Ancestral, mammilian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.*, **246**, 401–417.
19. Haber, J.E. (2000) Partners and pathways repairing a double-strand break. *Trends Genet.*, **16**, 259–264.
20. Löwer, R., Löwer, J. and Kurth, R. (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl Acad. Sci. USA*, **93**, 5177–5184.
21. Tristem, M. (2000) Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.*, **74**, 3715–3730.
22. Kjellman, C., Sjögren, H.-O. and Widegren, B. (1995) The Y chromosome: a graveyard for endogenous retroviruses. *Gene*, **161**, 163–170.
23. Sargent, C.A, Boucher, C.A., Kirsch, S., Brown, G., Weiss, B., Trundley, A., Burgoyne, P., Saut, N., Durand, C., Levy N. *et al.* (1999) The critical region of overlap defining the AZFa male infertility interval of proximal Yq contains three transcribed sequences. *J. Med. Genet.*, **36**, 670–677.
24. Foresta, C., Ferlin, A. and Moro, E. (2000) Deletion and expression analysis of *AZFa* genes on the human Y chromosome revealed a major role for *DBY* in male infertility. *Hum. Mol. Genet.*, **8**, 1161–1169.
25. Yen, P.H., Li, X.-M., Tsai, S.-P., Johnson, C., Mohandas, T. and Shapiro, L.J. (1990) Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. *Cell*, **61**, 603–610.
26. Chance, P.F., Abbas, N., Lensch, M.W., Pentao, L., Roa, B.B., Patel, P.I. and Lupski, J.R. (1994) Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum. Mol. Genet.*, **3**, 223–228.
27. Karsten, S.L., Lagerstedt, K., Carlberg, B.-M., Kleijer, W.J., Zaremba, J., Van Diggelen, O.P., Czartoryska, B., Pettersson, U. and Bondeson, M.-L. (1997) Two distinct deletions in the IDS gene and the gene W: a novel type of mutation associated with the Hunter syndrome. *Genomics*, **43**, 123–129.
28. Edelmann, L., Pandita, R.K. and Morrow, B.E. (1999) Low copy repeats mediate the common 3 Mb deletion in patients with velo-cardio-facial syndrome. *Am. J. Hum. Genet.*, **64**, 1076–1086.
29. Christian, S.L., Fantes, J.A., Mewborn, K., Huang, B. and Ledbetter, D.H. (1999) Large genomic duplicons map to sites of instability in the Prader–Willi/Angelman syndrome chromosome region (15q11–q13). *Hum. Mol. Genet.*, **6**, 1025–1037.
30. Peoples, E., Franke, Y., Wang, Y.-K., Pérez-Jurado, L., Paperna, T., Cisco, M. and Francke, U. (2000) A physical map, including a BAC/PAC clone contig of the Williams-Beuren syndrome-deletion region at 7q11.23. *Am. J. Hum. Genet.*, **66**, 47–68.
31. Dorschner, M.O., Sybert, V.P., Weaver, M., Pletcher, B.A. and Stephens, K. (2000) NF1 microdeletion breakpoints are clustered at flanking repetitive sequences. *Hum. Mol. Genet.*, **9**, 35–46.

32. Reiter, L.T., Hastings, P.J., Nelis, E., De Jonghe, P., Van Broeckhoven, C. and Lupski, J.R. (1998) Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am. J. Hum. Genet.*, **62**, 1023–1033.
33. Liskay, R.M. and Stachelek, J.L. (1986) Information transfer between duplicated chromosomal sequences in mammalian cells involves contiguous regions of DNA. *Proc. Natl Acad. Sci. USA*, **83**, 1802–1806.
34. Semino, O., Passarino, G., Brega, A., Fellous, M. and Santachiara-Benerecetti, A.S. (1996) A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am. J. Hum. Genet.*, **59**, 964–968.
35. Zerjal, T., Pandya, A., Santos, F.R., Adhikari, R., Tarazona, E., Kayser, M., Evgrafov, O., Singh, L., Thangaraj, K., Destro-Bisol, G. *et al.* (1999) The use of Y-chromosomal DNA variation to investigate population history. In Papiha, Deka and Chakraborty (eds), *Genomic Diversity: Applications in Human Population Genetics*. Kluwer Academic/Plenum Publishers, New York, NY, pp. 91–101.
36. Grimaldi, P., Scarponi, C., Rossi, P., March, M.R., Fabbri, A., Isidori, A., Spera, G., Krausz, C. and Geremia, R. (1998) Analysis of Yq microdeletions in infertile males by PCR and DNA hybridisation techniques. *Mol. Hum. Reprod.*, **4**, 1116–1121.
37. Vogt, P., Chandley, A.C., Hargreave, T.B., Keil, R., Ma, K. and Sharkey, A. (1992) Microdeletions in interval 6 of the Y chromosome of males with idiopathic sterility point to disruption of AZF, a human specific spermatogenesis gene. *Hum. Genet.*, **89**, 491–496.
38. Bishop, C., Guellaen, G., Geldwerth, D., Fellous, M. and Weissenbach, J. (1984) Extensive sequence homologies between Y and other human chromosomes. *J. Mol. Biol.*, **173**, 403–417.