AZF deletions and Y chromosomal haplogroups: history and update based on sequence

Peter H. Vogt

Section of Molecular Genetics & Infertility, Department of Gynecological Endocrinology & Reproductive Medicine, University of Heidelberg, Heidelberg, Germany

Address for correspondence: Voßstrasse 9, D-69115 Heidelberg, Germany. E-mail: peter_vogt@med.uni-heidelberg.de

AZF deletions are genomic deletions in the euchromatic part of the long arm of the human Y chromosome (Yq11) associated with azoospermia or severe oligozoospermia. Consequently, it can be assumed that these deletions remove Y chromosomal genes required for spermatogenesis. However, these ‘classical’ or ‘complete’ AZF deletions, AZFa, AZFb and AZFc, represent only a subset of rearrangements in Yq11. With the benefit of the Y chromosome sequence, more rearrangements (deletions, duplications, inversions) inside and outside the classical AZF deletion intervals have been elucidated and intra-chromosomal non-allelic homologous recombinations (NAHRs) of repetitive sequence blocks have been identified as their major cause. These include duplications in AZFa, AZFb and AZFc and the partial AZFb and AZFc deletions of which some were summarized under the pseudonym ‘gr/gr’ deletions. At least some of these rearrangements are associated with distinct Y chromosomal haplogroups and are present with similar frequencies in fertile and infertile men. This suggests a functional redundancy of the AZFb/AZFc multi-copy genes. Alternatively, the functional contribution(s) of these genes to human spermatogenesis might be different in men of different Y haplogroups. That raises the question whether, the frequency of Y haplogroups with different AZF gene contents in distinct human populations leads to a male fertility status that varies between populations or whether, the presence of the multiple Y haplogroups implies a balancing selection via genomic deletion/amplification mechanisms.

Key words: AZF deletions and amplicons in Yq11/Y chromosomal haplogroups and branches/Y sequence polymorphisms and AZF genetic redundancy/AZF gene content and male fertility status

Introduction

In 1976 Tiepolo and Zuffardi published the chromosome pictures of six men with a monocentric Yq1 chromosome all suffering from a severe impairment of spermatogenesis diagnosed in their testicular tissue sections (Tiepolo and Zuffardi, 1976). These Y chromosomal abnormalities were ‘de novo’ mutations, i.e., not observed in the Y chromosome of the patients’ father, thus no polymorphic events. The authors therefore, suggested that there must be a genetically functional AZoospermia factor (AZF) on the long arm of the human Y chromosome and since they did not assume that there is a spermatogenic function of the highly polymorphic heterochromatin domain in Yq12, they postulated that this AZF locus should be located in the distal part of the euchromatic Y long arm, i.e., in Yq11.23 (Tiepolo and Zuffardi, 1976). At that time there was no idea about the genetic content of this AZF locus. Protein encoding AZF genes, functionally expressed in human testis tissue, were not expected due to the extremely variable length of the Y chromosome in fertile men (Unnerus et al., 1967) including also the euchromatic Yq11 region (Soudek et al., 1973). However, this chromosome variability was stable in family pedigrees and able to assign individuals to specific populations. A functional active AZF chromatin domain was therefore first proposed, visible by the decondensation of the Y chromosome in the nuclei of spermatogonia before it pairs with the X chromosome forming a condensed X-Y chromatin structure in the spermatocyte nuclei (Speed et al., 1993; Vogt et al., 1995).

Today we know that there are at least fourteen protein encoding Y genes part of the AZF locus (Skaletsky et al., 2003; Vogt, 2005) and that the premeiotic pairing process of the sex chromosomes along the AZF chromatin is essential for a proper meiosis of the male germ cells (Armstrong et al., 1994; Turner et al., 2005). Moreover, with knowledge of the euchromatic Y sequence it is now also evident that the two faces of this chromosome, namely, being polymorphic on one site, being functional for spermatogenesis at the other site, are structurally intermingled, i.e., functional AZF genes are structurally linked to the Y specific highly polymorphic DNA regions. Accordingly, multiple
Y chromosomes have developed during human evolution distinguished now by a rooted pedigree of 153 Y chromosomal haplogroups (The Y Chromosome Consortium (YCC), 2002; Jobling and Tyler-Smith, 2003). Probably these Y structures were selected to drive the male reproductive fitness of the different human populations (Gill, 2002; Charlesworth, 2003). Molecular reasons of the dynamic Y sequence structure are mainly non-allelic homologous recombination (NAHR) of the locus-specific repetitive sequence block in distal Yp and Yq11 most of which are organised in large palindromes (Skaletsky et al., 2003).

This review aims to introduce the reader to the history of AZF deletions and the current knowledge of their association with different polyomorph rearrangements present in distinct Y haplogroups. The data presented are updated on the platform of the first genomic Y sequence now published in the data base (http://www.ensembl.org/Homo_sapiens/mapview?chr=Y) and on the first rooted pedigree of the human Y chromosomal haplogroups (The Y Chromosome Consortium (YCC), 2002; http://ycc.biosci.arizona.edu).

Mapping of AZF deletions in Yq11

The first successful approach narrowing the extension of the genomic region in Yq11 functional important for spermatogenesis (AZF region) was molecular deletion mapping. Using Y-specific DNA probes mapped in Yq11, breakpoints of genomic DNA samples of infertile men displaying a cytogenetically visible Y abnormality in Yq11 were ordered in linear interval maps (Affara et al., 1986; Vergnaud et al., 1986; Oosthuizen et al., 1990; Bardoni et al., 1991; Ma et al., 1992; O’Reilly et al., 1992). Sequence analyses of numerous Y clones exhibited extensive sequence homologies between the Y and other human chromosomes especially to the X chromosome (Bishop et al., 1984). Thus restriction fragment length polymorphism (RFLP) probes were needed in order to mark the Y located restriction fragments unambiguously.

The different Yq11 abnormalities observed in infertile men and used for these first AZF mapping studies were cytogenetically visible and are summarized schematically in Figure 1. Studying the testicular histology of these Yq- patients, it was suggested that AZF most probably has a premeiotic and postmeiotic spermatogenesis function (Vogt et al., 1993). Low numbers of spermatagonia besides a normal number of Sertoli cells were most often reported in all cases with terminal Yq11 deletions and unbalanced Yq translocations (Tiepolo and Zuffardi, 1976; Yunis et al., 1977; Andersson et al., 1988, Hartung et al., 1988; Bardoni et al., 1991) whereas, the testicular histology of infertile men with a balanced Yq translocation were less severe. When infertile, their testicular histology showed disruption of spermatogenesis at meiosis or after the formation of spermatids (Faed et al., 1982).

A large number of men with a gross Yq11 abnormality was not detectable by the routine Giemsa-staining protocol. In these cases the broken Yq11- chromosomes of two cell nuclei have fused together in a secondary meiotic or mitotic rearrangement forming a di-centric Yq11 chromosome (‘dic-(Yp)’) with the duplicated Y short arms (case B in Figure 1). This Yq11 rearrangement is frequently accompanied with loss of the Yq11 chromosome in some nuclei leading to a mosaic 45,X0/46,X,dic(Yp) karotype. Even today, the dic-(Yp) chromosomes are often wrongly described as being ‘normal’ because of the size similar to that of normal Y chromosomes (Siffroi et al., 2000). However, both can easily be distinguished after staining the chromosomes with quinacrine. All dic-(Yp) chromosomes have lost the large fluorescent heterochromatin block in Yq12 marking the normal Y chromosome in metaphase and interphase nuclei. Only the normal Y is fluorescent, the dic-(Yp) chromosomes are always non-fluorescent (‘nf’) and are therefore called ‘Ynf’ chromosomes (Sandberg, 1985). Today Ynf

Figure 1. Schematic view on seven typical chromosomal rearrangements in Yq11 (A–G) which are associated with the occurrence of azoospermia or oligozoospermia because of disruption of the functional structure of the AZoospermia Factor (AZF). (A) Monocentric Yq11- chromosomes are distinguished from the normal Y chromosome by quinacrine staining or by FISH with a heterochromatic DNA probe (DYZ1; DYZ2). The same is possible with the dicentric dic-Yp chromosomes (B) and ring-Y chromosomes (C) which are often found together with a 45,X0 cell line in the patient’s lymphocyte nuclei. After breakage in Yq11 and loss of the Yq11- chromosome the distal Yq11 part can also translocate to the short arm of a second Y chromosome (D), or to the short arm of the X chromosome (F). The stable Yq11- chromosome (A) can be also translocated to an autosome in a second rearrangement (E). Translocations of the distal Yq11 part to autosomes (G) are mainly to the acrocentric chromosomes: 14, 15, 21, 22 and also found in fertile men dependent on the Yq11 breakage site.
chromosomes can be easily diagnosed by fluorescence in-situ hybridization (FISH) using a centromeric Y specific DNA probe of the DYZ3 locus (Genome Data Base accession no. 1665101). Diagnostic is the presence of two and sometimes four FISH spots on the patients’ metaphase Y chromosomes (Vogt, 2005).

There is an inherent instability of the whole chromosome structure in Yq11 recognised by different quantities of X0 cells (1−80%) in the patient’s lymphocytes and sometimes interstitial deletions in the Y short or long arm (Köhler and Vogt, 1994). The X0 cells are especially diagnosed in the lymphocytes of infertile men with a Ynf or Y-ring chromosome (Fryns et al., 1978; Diekmann et al., 1992; Henegarui et al., 1997). Therefore, the presence of a Ynf chromosome is sometimes associated with additional pathologies: undescribed testis and ambiguous genitalia (Thangaraj et al., 2003) indicating the 45, X0 cells also in the patients’ gonad tissues. Similar pathologies were also described from patients with an AZFc deletion and a 45, X0 cell line (Siffroi et al., 2000; Jaruzelska et al., 2001; Papadimas et al., 2001). It suggests that Ynf chromosomes might be one of the final rearrangements of a generally unstable Y chromosome after the occurrence of a classical (i.e., complete) AZFc deletion (Vogt, 2004).

The positive identification of a Ynf chromosome in the patient’s lymphocyte nuclei can explain the occurrence of azoospermiain his testis tubules because the meiotic pairing of both Yp arms (Chandler et al., 1986) disrupts the pre-meiotic X–Y pairing process and meiotic silencing of the unpaired X and Y chromosomes at pachytene is inhibited (Baarends et al., 2005; Turner et al., 2005). Indeed, in all cases where a testicular biopsy was evaluated in this patient group spermatogenesis was blocked before or at meiosis (for review see: Vogt, 1996). Since the nature and extent of the X–Y pairing structure at the meiotic prophase temporarily involves the entire euchromatic Y long arm (Ashley, 1984; Chandler et al., 1984) premeiotic ‘X–Y pairing sites’ in Yq11 are suggested. All cytogenetically visible Y chromosomal rearrangements in Yq11 would then interfere with the premeiotic X–Y pairing process because of the already premeiotic silencing of the unsynapsed X chromatin in the leptotene phase (Baarends et al., 2005; Turner et al., 2005). Consequently a chromosomally based AZF spermatogenesis function (i.e., active premeiotic ‘X–Y chromosomal pairing sites’ in Yq11) would be able to explain easily all cases of infertile men with cytogenetically visible Yq deletions; a meiotic disruption of spermatogenesis would be the expected main result.

Besides this proposed ‘AZF-chromatin function’, molecular identification of the first putative AZF genes, RBMY (Ma et al., 1993) and DAZ (Reijo et al., 1995), indicated that the human Y chromosome is also encoding essential protein encoding Y genes functional for spermatogenesis. At least three AZF genes should be present in Yq11 functional at different phases of the spermatogenic cycle. This was concluded from the results of a large screening program for putative AZF micro-deletions (i.e., not visible in the microscope) in 370 infertile men with a normal karyotype (46,XY) mapping three different ‘de novo’ deletions in Yq11 in 12 men with azoospermia or severe oligozoospermia (Vogt et al., 1996). The three AZF deletion intervals designated as AZFa, AZFb, and AZFc sequence regions were confirmed in multiple similar studies (Vogt, 1998; Krausz et al., 2003). These now also called ‘classical’ or ‘complete’ AZF micro-deletions are caused by intrachromosomal recombinations between homologous repetitive sequence blocks (AZFa: Blanco et al., 2000; Kamp et al., 2000; Sun et al., 2000; AZFc: Kuroda-Kawaguchi et al., 2001; AZFb: Repping et al., 2002) and are always associated with the occurrence of a distinct testicular pathology (Krausz et al., 2000; Krausz et al., 2003; Vogt, 2005). Today we know that there are not only 3 but at least 14 protein encoding Y genes in these AZF regions (Table I).

Polymorphic Y-fragments in AZF regions establish first Y chromosomal haplogroups

If in a family pedigree, a Y-specific DNA probe detects a variant banding pattern in genomic DNA blots, it is found often to be stably inherited from father to son. This ‘polymorphic’ stability is probably based on the lack of regular interchromosomal recombination events (crossing overs) along the complete male specific Y sequence. Polymorphic Y sequence variants became therefore quickly established to design ‘compound haplotypes’ and Y chromosomal haplogroups for tracing the evolutionarily relatedness of the present human populations (Torroni et al., 1990; Jobling et al., 1996; Semino et al., 1996). The first polymorphic sequence variants were observed with Y probes now known to map to the three ‘classical’ AZF deletion intervals (Figure 2).

The DYS11 DNA locus (12f2 probe; Casanova et al., 1985) has been mapped to the distal HERV15 sequence block of the AZFa deletion in proximal Yq11 (Blanco et al., 2000; Kamp et al., 2000; Sun et al., 2000). Its repetitive hybridisation pattern on genomic female DNA blots indicated that 12f2 has many related sequences on the X chromosome and on autosomes. However, the two variable fragment lengths (8, 10.4 kb) observed after TaqI restriction were only present in genomic male DNA, indicating their Y chromosomal origin (Figure 2). Analysis of the frequency of this 2.4 kb deletion polymorphism (here designated as DYS11 12f2-2.4 kb allele) in different populations from Europe, Africa and Asia revealed its specificity for Caucasian populations because absent in African Blacks, in Orientals and in Native Americans (Semino et al., 1996). Most interesting, the frequency of the 12f2-2.4 kb deletion decreased from the Near-East to northwestern Europe populations reflecting the neolithic demic diffusion of the ancient farming cultures. Today we know that the 12f2-2.4 kb deletion is the derived state (12f2-B allele) of the undeleted ancient DYS11 sequence (12f2-A allele) and that this deletion must have occurred at least two times during evolution of the human population history (Blanco et al., 2000).

The polymorphic 50f2 DNA probe (Genome Data Base accession code: DYS7: 168024) cross hybridised to five specific Y-fragments (A–E), after EcoRI restriction (Figure 2). 50f2/A + B was mapped to proximal Yp, 50f2/C to AZFc, 50f2/D to the Y centromer and 50f2/E to AZFb (Vogt et al., 1996). The molecular base of length variation of the 50f2/B fragment in Yp is a mini-satellite block (MSY1) containing 48–114 copies of an AT rich 25 nt. long sequence unit (Jobling et al., 1998). With a virtual heterozygosity of 99.9%, MSY1 is by far the most variable DNA locus on the human
Table I. Human Y genes with putative spermatogenesis function mapped to the AZFa, AZFb, AZFc deletion intervals

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Number of copies and code</th>
<th>Protein homolog to</th>
<th>Tissue RNA expression</th>
<th>Copies in Yp interval</th>
<th>Location in Yq11</th>
<th>X chromosome homolog</th>
<th>Autosome homolog</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPY2</td>
<td>Basic Protein Y 2</td>
<td>BPY2.1 – 3</td>
<td>Novel</td>
<td>Only testis</td>
<td>no</td>
<td>AZFc</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>CDY1</td>
<td>Chromo Domain Y1/2</td>
<td>CDY1.1–2</td>
<td>Chromatin-Protein and histone-acetyltransferases</td>
<td>Only testis only testis</td>
<td>no</td>
<td>AZFb + Yq11-D11 (CDY2) AZFc (CDY1)</td>
<td>no</td>
<td>6p24; CDYL</td>
</tr>
<tr>
<td>CDY2</td>
<td></td>
<td>CDY2.1–2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSPG4LY</td>
<td>Chondroitin sulfate proteoglycan 4 Like Y</td>
<td>CSPG4LY.1 CSPG4LY.2</td>
<td>Cadherins</td>
<td>Only testis</td>
<td>no</td>
<td>AZFc</td>
<td>no</td>
<td>15q24; CSPG4</td>
</tr>
<tr>
<td>DAZ</td>
<td>Deleted in Azoospermia</td>
<td>DAZ1, DAZ2, DAZ3, DAZ4</td>
<td>RNA binding RRM proteins</td>
<td>Only testis</td>
<td>no</td>
<td>AZFc</td>
<td>no</td>
<td>3p24; DAZL 2q33; BOULE</td>
</tr>
<tr>
<td>DBY aka DDX3Y</td>
<td>DEAD Box Y</td>
<td>1</td>
<td>DEAD box RNA helicases</td>
<td>Multiple</td>
<td>no</td>
<td>DBX aka DDX3X</td>
<td></td>
<td>no</td>
</tr>
<tr>
<td>EIF1AY</td>
<td>Essential Initiation Transl. Factor 1A Y</td>
<td>1</td>
<td>Translation Initiation Factor</td>
<td>Multiple</td>
<td>no</td>
<td>AZFb</td>
<td>EIF1AX</td>
<td>no</td>
</tr>
<tr>
<td>GOLGA2LY</td>
<td>Golgi autoantigen, golgin Subfamily a2 Like Y</td>
<td>GOLGA2LY.1 GOLGA2LY.2</td>
<td>CIS GOLGI Matrix Protein GM130</td>
<td>Only testis</td>
<td>no</td>
<td>AZFc</td>
<td>no</td>
<td>9q34; GOLGA2</td>
</tr>
<tr>
<td>HSFY</td>
<td>Heat-Shock transcription Factor Y linked</td>
<td>HSFY.1–2</td>
<td>HSP-2 like</td>
<td>Testis, kidney</td>
<td>no</td>
<td>AZFb</td>
<td>no</td>
<td>6q22; HSP2</td>
</tr>
<tr>
<td>PRY</td>
<td>PTP - BL Related Y</td>
<td>PRY.1–2</td>
<td>Protein tyrosine Phosphatase</td>
<td>Only testis</td>
<td>prox. Yp11 pseudogenes</td>
<td>AZFb AZFc; pseudogenes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>RBMY</td>
<td>RNA Binding Motif Y-linked</td>
<td>RBMY1.1–6</td>
<td>RNA binding RRM - Proteins</td>
<td>Only testis</td>
<td>prox. Yp11 pseudogenes</td>
<td>AZFb AZFc; pseudogenes</td>
<td>RBMX</td>
<td>HNRP G-T retrogene</td>
</tr>
<tr>
<td>RPS4Y2</td>
<td>Ribosomal Protein S4 Y linked 2</td>
<td>1</td>
<td>S4 ribosomal protein</td>
<td>Multiple</td>
<td>distal Yp11</td>
<td>AZFb</td>
<td>RPS4X</td>
<td>no</td>
</tr>
<tr>
<td>SMCY</td>
<td>Selected Mouse C DNA Y</td>
<td>1</td>
<td>H-Y antigen HLA B7</td>
<td>Multiple</td>
<td>No</td>
<td>AZFb</td>
<td>SMCX</td>
<td>no</td>
</tr>
<tr>
<td>USP9Y</td>
<td>Ubiquitin specific proteasease 9 Y</td>
<td>1</td>
<td>Ubiquitin-specific proteasease</td>
<td>Multiple</td>
<td>No</td>
<td>AZFb</td>
<td>USP9X aka DFFRX</td>
<td>no</td>
</tr>
<tr>
<td>XKRY</td>
<td>X - Kell blood group precursor related Y</td>
<td>XKRY.1–2</td>
<td>Putative membrane transport protein</td>
<td>Only testis</td>
<td>No</td>
<td>AZFb + Yq11-D11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---


b Additional RNA populations with smaller lengths were found only in testis tissue (Lahn and Page, 1997; Ditton et al. 2004).
Y chromosome and therefore also useful for dating paternal lineages and for forensic studies (Jobling et al., 1999).

After analyses of 859 men from 46 different populations absence of the 50f2/C fragment was found in 55 of them and duplications were found in 8 men (Jobling et al., 1996). Additionally an independent length variation between 7.5–8.5 kb was found for the 50f2/B fragment in different individuals (Jobling et al., 1998; see also Figure 2). With knowledge of the Y-sequence, 50f2/C was mapped in the u3-marked sequence block of proximal AZF. The 50f2/E fragment was mapped proximal to it with a distance of 1.4 Mb in AZFb (Figure 3). Since some 50f2/C deletions were associated with deletions of neighboured RBMY gene copies (Jobling et al., 1996), these polymorphic AZF deletions should include at least part of the AZFc t-amplicons and the b2 amplicon, i.e., would have an extension of ~500 kb (Figure 3). However, most polymorphic 50f2/C deletions were reported to be smaller not including other known DNA loci (Jobling et al., 1996). The frequencies of 50f2/C deletions are variable in the 12 human populations analysed and taking their different Y chromosomal haplogroups into account, six independent deletion events and four duplication events were identified (Jobling et al., 1996). Thus multiple founder rearrangements must have occurred during human evolution in the proximal AZFc sequence region which include all the 50f2/C EcoRI fragment. The highest frequency (55%) of 50f2/C deletions was found in Fins (11 of 20 men had this deletion) making it unlikely that this deletion was associated with spermatogenic failure at least in this population.

The third polymorphic AZF-locus (DYS1; 49f probe; Lucotte and Ngo, 1985) is composed of 18 TaqI restricted genomic Y-DNA fragments (A–R; Figure 2) and part of the DAZ gene structure in AZFc (Figure 3A). The variability of six TaqI fragments (A, C, D, F and I: present or absent; A and D modified in length: A1 = 11.1 kb; A2 = 14.8 kb; A3 = 17.3 kb; A4 = 19.6 kb; D1 = 7.2 kb; D2 = 6.9 kb) seemed to occur independently from each other and therefore established the first multi-allelic Y chromosome marker system (Ngo et al., 1986). The DYS1 length variabilities are reflecting the variable number and sequence variants of the repetitive exon 7 copies in the four DAZ genes (also called DAZ-repeats) to which the 49f probe cross hybridised; the cross hybridising K and L female fragments belong to the autosomal DAZL gene copy on the short arm of chromosome 3 (Saxena et al., 1996; Vogt et al., 1997).
Figure 3. Schematic view on some possible variations of the AZFb/c amplicon structure in distal Yq11. (A) Structural organisation of the different amplicons in the Y chromosome of the R* haplogroup in five palindromic structures (P1–P5). The amplicons’ colour code and nomenclature is derived from the work of Skaletsky et al. (2003). The polymorphic 49f-sites (DYS1 DNA locus) in the DAZ exon 7 repeats and the polymorphic Yfm1 marker, 25–30 kb distal to the DAZ genes (Ewis et al., 2002), are located in the red amplicons. The polymorphic 50f2-DNA-locus (DYS7) is located in the AZFc-u3 region and marked with the putative extensions of its deletions in grey colour. Some 50f2/C deletions include the deletion of neighboured RBMY gene copies, some the deletion of the 50f2/E sequence site, i.e., have a size of up to 4 Mb (Jobling et al., 1996; Jobling and Tyler-Smith, 2003). The polymorphic YAP deletion site (+/-) is marked distal to the P4 amplicon in AZFb. The location of the 12f2-2-kb deletion polymorphism in distal HERV15yq of the AZFa deletion interval is marked in Yq11.21. STS markers selected for the detection of partial and total AZFb/c deletions in the Y–R* chromosome (Repping et al., 2003) are given in black above the AZFb/c amplicon structure. (B) Schematic view on the extensions of possible partial AZFb/c deletions in the Y–R* chromosome. All are based on the assumption that homologous amplicons (same colour) with the same polarity are frequently recombining with a NAHR mechanism and that these deletions are the subsequent mutation events (Yen, 2001). Accordingly, the gr/gr deletions designated by Repping et al. (2003) are split into the three subgroups: g1/g2, r1/r3, r2/r4. Please note that the STS deletion pattern marking the different partial AZFb/c deletions are associated with the Y sequence of the R* haplogroup and might be different in the Y chromosome of the other Y haplogroups. (C) Schematic overview on the putative 13 ampliconic AZFb/c inversion events in the Y–R* chromosome. To decrease the complexity of the picture the drawings for two further inversions (b1 ↔ b4; grey1 ↔ grey2) have been omitted. (D) Schematic view on the extensions of the so-called ‘complete’ or ‘classical’ AZFb and AZFc deletions and their overlap in the Y–R* chromosome. The two different breakage-fusion sites in the P5 palindrome (I and II) are indicated in the work of Repping et al. (2002).
Formal proof that the variable DAZ gene structures in exon 7 indeed form the molecular base for the polymorphic DYS1-49f blot fragments, the TaqI restriction patterns for each DAZ gene copy of the RP11-donor was extracted ‘in silico’ from the corresponding DAZ-BAC sequence data (Fernandes et al., 2002; Jovelin et al., 2003). The 10 kb long B was found in the DAZ1 and DAZ2 gene copy, the 7.15 kb long D2 and 3.1 kb long I fragments only in DAZ2, the F fragment (5 kb) and the A4 fragment (19.6 kb) in the DAZ3 gene copy and the A2 fragment (14.8 kb) only in the DAZ4 gene exon 7 repeat. No TaqI fragment with length of C (7.9 kb) could be identified in the DAZ locus of the RP-11 Y chromosome (Table II). According to the nomenclature of Torroni et al. (1990), the corresponding DYS1 haplotype for the RP-11 donor should be then a variant of haplotype ‘VIII’ with the polymorphic TaqI fragments: A2, C0, D1, F1, and I1. This Y chromosomal haplotype has been reported with a median frequency in Caucasian male populations (5.6%) and absent in Africans (Torroni et al., 1990). Although one origin of the polymorphic DYS1-TaqI fragments might be frequent single nucleotide mutations in DAZ exon 7 TaqI restriction sites (Jovelin et al., 2003), additionally also real DAZ gene deletions and duplications should modify the TaqI restriction patterns of the DYS1-DAZ gene locus as well. They may include only the DAZ-repeat as described earlier (Vogt and Fernandes, 2003) or complete DAZ gene copies as described below.

A substantial number of additional polymorphic marker fragments in the AZFb and AZFc deletion intervals were found in recent years in different human populations (AZFb: LLY22g, 92R7, AZFc: poxY1), identified and studied first by blot experiments (Mathias et al., 1994; Jobling et al., 1996; Oakey and Tyler-Smith, 1990) then by the more convenient PCR format (Jobling and Tyler-Smith, 1995; Rossler et al., 2000). One of the most popular Y population marker became the YAP element (Genome Data Base accession code: DYS287: 196899) with a unique origin and heterogenous frequency in different human populations (Hammer, 1995). YAP is the insertion of a short interspersed nucleotide element (SINE) ‘Alu-repeat’ sequence in Yq11. Most of these highly repetitive sequence elements have an AluI restriction site in their 300 nucleotide long sequence unit and are therefore also called ‘Alu-repeats’ (Deininger et al., 1981). The YAP insertion could be mapped to the proximal region of AZFb in the RP-11 BAC 169D1 sequence (GenBank accession no. AC010137) distal to the P4 palindrome (Figure 3A; S. Kirsch, personal commun.). YAP + chromosomes are frequently present in the Japanese (42%) and most other Asian populations although absent in the Taiwanese and with the highest frequency in the sub-Saharan African populations (Hammer and Horai, 1995).

Since all the described polymorphic sequence variants in the different AZF regions are broadly distributed in different human populations no reduced fertility is expected to be associated with any of them although there might be exceptions. A low fertility in men with haplogroup Y-hg26+ (now K* (xP), see below) was reported in a population of Denmark (Krausz et al., 2001) and in a Japanese men population the occurrence of azoospermia

### Table II. DYS1-49f TaqI restriction fragments in DAZ locus of RP11 Y chromosome

<table>
<thead>
<tr>
<th>DYS1 locus letter code</th>
<th>DAZ1</th>
<th>DAZ2</th>
<th>DAZ3</th>
<th>DAZ4</th>
<th>in silico 49f % homology</th>
</tr>
</thead>
<tbody>
<tr>
<td>A4</td>
<td>19612</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>81</td>
</tr>
<tr>
<td>A3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>–</td>
<td>–</td>
<td>14853</td>
<td>–</td>
<td>&gt;80%</td>
</tr>
<tr>
<td>A1</td>
<td>10094</td>
<td>10100</td>
<td>–</td>
<td>–</td>
<td>81</td>
</tr>
<tr>
<td>B</td>
<td>7152</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>81</td>
</tr>
<tr>
<td>C</td>
<td>3112</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>81</td>
</tr>
<tr>
<td>D1</td>
<td>5854</td>
<td>–</td>
<td>5008</td>
<td>5857 (2)f</td>
<td>97/83</td>
</tr>
<tr>
<td>D2</td>
<td>4782</td>
<td>4781</td>
<td>–</td>
<td>4783</td>
<td>81</td>
</tr>
<tr>
<td>E</td>
<td>4020</td>
<td>4020</td>
<td>–</td>
<td>4021</td>
<td>83</td>
</tr>
<tr>
<td>F1</td>
<td>2975</td>
<td>2976</td>
<td>2975</td>
<td>2975</td>
<td>83</td>
</tr>
<tr>
<td>H</td>
<td>1877</td>
<td>1876</td>
<td>1876</td>
<td>1876</td>
<td>99</td>
</tr>
<tr>
<td>I</td>
<td>2076</td>
<td>2076</td>
<td>2075</td>
<td>2075</td>
<td>81</td>
</tr>
<tr>
<td>J</td>
<td>1419 (3)</td>
<td>1419 (3)</td>
<td>1419 (2)</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>K (DAZL)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>83</td>
</tr>
<tr>
<td>L (DAZL)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>86</td>
</tr>
<tr>
<td>M</td>
<td>1237</td>
<td>1236</td>
<td>1236</td>
<td>1236</td>
<td>81</td>
</tr>
<tr>
<td>N</td>
<td>966</td>
<td>966</td>
<td>966</td>
<td>966</td>
<td>80</td>
</tr>
</tbody>
</table>

*All genomic fragments present in den RP-11 donors DAZ-BAC clone sequences and with a ‘in silico’ >80% homology to the 49f sequence (GenBank accession no.:AF414183) were listed and associated with the letter code of Ngo et al. (1996) in the first column (see also Figure 2). Italic letters mark the polymorphic fragments.

b ‘–’ indicate that these polymorphic fragment lengths are absent in the RP-11 DAZ-BAC sequences.

Numbers in parentheses indicate that this fragment length is found with 2, 3,… copies.
seems to be associated with distinct haplotypes of the polymorphic Yf1 marker (Kuroki et al., 1999; Ewis et al., 2002) mapped distal to the DAZ gene copies in AZFc (Figure 3A).

Y chromosomal haplogroups are rooted in one complex pedigree

Since it was attractive to combine these highly informative sequence variants in ‘compound’ haplotypes structured in an hierarchical sequential order (Jobling and Tyler-Smith, 1995; Hurles and Jobling, 2001) more sequence variants of the Y chromosome were searched for extensively in the last years. However, only after it became possible to visualize sequence variants directly by specific heteroduplex formations in automated denaturing high performance liquid chromatography (DHPLC) experiments (Underhill et al., 2000), the number of compound haplotypes became quickly expanding and the construction of a first comprehensive Y haplogroup pedigree became possible (Figure 4).

A detailed description of the 153 Y chromosomal haplogroups is beyond the scope of this review and the reader is advised to read the original papers (The Y Chromosome Consortium (YCC), 2002; Jobling and Tyler-Smith, 2003), respectively, to visit the corresponding websites (http://ycc.biosci.arizona.edu). The YCC nomenclature system marks the different Y chromosomal compound haplogroups (branches: A–R) in a unique hierarchical order. Analysis of a distinct marker set from this pedigree can be used to identify the haplogroup of each given Y DNA sequence (Paracchini et al., 2002; Sanchez et al., 2003; Brion et al., 2005).

The published human Y DNA sequence belongs to Y chromosomal haplogroup R1*

The male specific region (MSY) of the Y sequence of the RP-11 donor (35 Mb) has now been published (Skaletsky et al., 2003) and deposited in different databases (GenBank: http://www.ncbi. gov; ENSEMBL: http://www.ensembl.org/Homo_sapiens/mapview?chr = Y). It is a mosaic of X-transposed, X-degenerated and repetitive Y-specific sequence blocks. The sequence is the result of a merged contig of bacterial artificial chromosome (BAC) clones each containing a portion of the MSY from the same individual, the RP-11 donor, an anonymous man of the result of a merged contig of bacterial artificial chromosome 11 donor encompassing at least three distinct sequence species, the Y long arm (Yq12) was estimated with The polymorphic heterochromatic Y region in the distal part of the Y long arm (Yq12) was estimated with 40 Mb in the RP-11 donor encompassing at least three distinct sequence species, DYZ1, DYZ2, DYZ18 each of which forms long and homogenous tandem arrays (Skaletsky et al., 2003). An interstitial heterochromatic repetitive sequence block of approximately 400 kb comprising > 3000 tandem repeats of a 125 nt. sequence unit (DYZ19) interrupts the distal large X-degenerate sequence block in AZFb (Figure 3).

No sequence informations were given for the centromeric repetitive alphoid DNA sequence block (~1 Mb). The pericentromeric region (~2 Mb) flanking this block in proximal Yq11 (Kirsch et al., 2004) was also not included in this Y sequence. It was revealed only recently and is composed by segmental duplications of sequence blocks also located in the pericentromeric region of different autosomes, namely chromosomes 1, 2, 3, 19, 16, and 22 (Kirsch et al., 2005). The pericentromeric Y chromosomal sequence region has therefore evolved more by inter-chromosomal than intra-chromosomal duplication events which have evolved the large repetitive sequence blocks (‘amplicons’) in distal Yq11 (Skaletsky et al., 2003).

It has long been predicted that especially the Y long arm in Yq11 is composed of numerous Y-specific repetitive sequence blocks (Foote et al., 1992; Kirsch et al., 1996). Sequence analysis has now confirmed this assumption. However, beyond that it has revealed the unique pattern of large repetitive sequence blocks (amplicons) ranging in length between 115 kb and 678 kb in distal Yq11 (Figure 3). Most interesting, these amplicons were structurally assembled in palindrome structures in AZFb and AZFc (P1: 2.9 Mb; P2: 246 kb; P3: 736 kb; P4: 419 kb; P5: 996 kb) (Figure 3) and between the AZFb and AZFf deletion intervals (P6: 266 kb; P7: 30 kb; P8: 75 kb), respectively. Their arms are highly symmetrical and comprise ~25% of the complete Y-specific sequence class, that is 5.7 Mb of the genomic Yq11 sequence.

Sequence analyses of homologous palindromic arms revealed extensive homologies between 99.94–99.997% along the complete amplicon sequence (Kuroda-Kawaguchi et al., 2001; Repping et al., 2002; Skaletsky et al., 2003). This suggests that the functional integrity of the Y genes mapped in the ampliconic AZF sequence regions is maintained by frequent gene conversions—that is non-reciprocal transfer of sequence information—between the homologous palindromic arms (Rozen et al., 2003).

The molecular extension of the three ‘classical’ AZF intervals in Yq11 is roughly ~8.73 Mb, that is ~60% of the complete Yq11 euchromatic sequence region. The AZFb interval overlaps with the proximal part of the AZFc deletion interval (Figure 3D) as was first proposed by a genomie YAC contig analysis (Kirsch et al., 1996).

Considering the derived state of the marker M207 in position 139.206 in the RPCL-11 BAC clone 386L3 (GenBank accession no. AC006376) it can be deduced that the Y sequence derived from the RP-11 donor belongs to Y haplogroup R* (Figure 4). This finding could be confirmed and further specified to

Figure 4. Schematic view of the phylogenetic tree of the 153 binary Y chromosomal haplogroups based on the references The Y chromosome Consortium (YCC) (2002) and Jobling and Tyler-Smith (2003) with modification of the colour code for better reading and distinguishing all Y haplogroups. The large letters A–R and Y at the left symbolise the main branches (clades) of the Y phylogeny. Along the horizontal pedigree-lines all bilallelic markers used for distinguishing the branches and subsequent haplogroups are given in blue colour. The nomenclature of the haplogroups for all branches is given at the right in the colour of the corresponding Y-branch. Haplogroups marked with an asterik (R*) are ‘paragroups’ meaning that they are not further defined by a derived marker. More details of the markers, together with further information about the nomenclature rules can be found at the YCC website (http://ycc.biosci.arizona.edu).
Y haplogroup R1* by a biallelic single nucleotide variant (SNV) marker in the yellow amplicon of the AZFc region (Fernandes et al., 2004).

Assuming a mutation rate of $1.6 \times 10^{-9}$ per nucleotide per year and $2.2 \times 10^{-8}$ conversions per duplicate nucleotide per generation (i.e., per 20 years) it has been calculated that along the 5.4 Mb length of the eight Y palindromes (i.e., $2.7 \times 10^8$ duplicated nucleotides) on average 600 duplicated nucleotides would have undergone arm-to-arm gene conversion, thus distinguishing the Y chromosome of father and son (Rozen et al., 2003). If this holds true, it must be concluded that in each of the 153 Y haplogroups identified (Figure 4) there should be a distinct Y reference sequence which is different from that of the RP-11 donor belonging to Y haplogroup R1*. Beyond of that the sequence variants in the highly polymorphic MSY1 locus and of the ~280 Y microsatellites (the so called Y STR-loci) are further specifying the Y sequences in each haplogroup establishing specific Y-lineages (Kayser et al., 2000). These STR variants are also spread in the three AZF subintervals in Yq11 but not yet mapped precisely in comparison to the STS/SNV maps in the same intervals (Kayser et al., 2004).

Mapping of AZF deletions in men of different Y haplogroups

The frequency of classical AZFa, AZFb and AZFc deletions with a molecular extension first defined by molecular deletion mapping (Vogt et al., 1996), then by the presence and absence of distinct border STS markers (AZFa: Kamp et al., 2001; AZFc: Kuroda-Kawaguchi et al., 2001; AZFb: Repping et al., 2002) seems not to be variable in the different chromosomal haplogroup (Paracchini et al., 2000; Quintana-Murci et al., 2001b; McElreavey and Quintana-Murci, 2003; Carvalho et al., 2003, 2004), although the data of Blanco et al. (2000) and Paracchini et al. (2000) suggests that this might not be true for AZFa deletions. Whether the different proximal border lines of the ‘classical’ AZFb deletions (Figure 3D) are associated with some distinct Y haplogroups is not yet known. However, it can be safely assumed that the number of amplicons and palindromes in the structure of the AZFa and AZFc regions as shown in Figure 3 and associated with Y haplogroup R* is probably variable in at least some of the other Y haplogroups and a large number of distinct ‘partial’ (to distinguish them from the molecular extensions of the classical complete AZFb/c deletions) AZFb and AZFc deletions would be the natural consequence (Yen, 2001; Hurles and Jobling, 2003). The NAHR based mechanism causing the intra-chromosomal recombinations and subsequent deletions in the ampliconic AZF sequence region is based on the same polarity of the homologous amplicons. The occurrence of b1/b3, g1/g2, r1/r3 and r2/r4 recombinations in the R*-Y-sequence resulting in different ‘partial’ AZFb/c deletions were therefore already predicted some years ago (Yen, 2001).

First partial AZF deletion studies pointed to a different number and structure of the DAZ genes in the red amplicons of the AZFc interval (Moro et al., 2000; de Vries et al., 2002, Fernandes et al., 2002; Ferlin et al., 2002, 2004; Ferras et al., 2004). However, the analyses of the corresponding AZFc amplicon structures did not include an analysis of the associated Y chromosomal haplogroups and sometimes were restricted to only a qualitative DAZ-STS or -SNV deletion analysis. Considering the high rate of gene conversions in the same palindromic sequence regions (see previous chapter) the conclusion of an associated ‘real DAZ gene deletion’ drawn from solely a pattern of STS/SNV deletions was therefore probably sometimes wrong. Because of the inherent instability of the location of single nucleotide marker sites in the palindromic amplicons (Rozen et al., 2003), the presence of a real partial AZF deletion needs to be generally confirmed by supporting experiments like FIBER-FISH (Repping et al., 2003), or specific DNA blot experiments (Fernandes et al., 2002), or similar quantitative gene copy deletion assays (Bienvenu et al., 2001; Macchev et al., 2004). Additionally, if judged as being a clinically significant putative causative agent for the man’s infertility, the analysis of a DNA sample from the patient’s father (or brother) and identification of the family’s Y chromosomal haplogroup is strongly recommended.

Due to the generally high dynamic palindromic sequence structure in the ampliconic AZFb and AZFc sequence regions and the presence of similar albeit smaller repetitive sequence blocks along the whole sequence in Yq11 (Skaletsky et al., 2003) multiple genomic rearrangements causing partial deletions in the ‘classical’ AZF sequence regions are expected. Most of them might occur frequently in the different lineages of the human Y chromosome currently present in the global population of over three billion men. Methods which compare the age of the Y-lineage displaying a distinct variant of the R1*-AZFb and AZFc amplicon structure would then help to determine whether the structure identified is compatible with neutrality (i.e., fertility) or associated with some spermatogenic failure effects (Sabeti et al., 2002). Thus, the preference of multi-copy genes especially in AZFb and AZFc can reflect genetic redundancy, but also some functional constraints from the germ line as for example representing a counterbalance for the unstable AZFb/c amplicon structure by reducing the risk of male infertility via a continuous genomic deletion/amplification mechanism in these AZF regions.

The first systematic screen for partial AZF deletions associated with specific Y chromosomal haplogroups was performed by Repping et al. (2003). In a study with some selected STS loci mapped in the AZFb/c amplicon structure of the R*-Y-sequence (Figure 3), 22 partial AZFb deletions were identified in a screen of 689 individuals. These genomic deletions were confirmed by FIBER-FISH and summarized under the pseudonym ‘gr/gr’ deletions because they displayed the same deletion pattern of the green and red FISH signals in the men’s lymphocyte nuclei (Repping et al., 2003). Another partial AZF deletion probably based on a b1/b3 recombination event in distal AZFb (Figure 3) was found in one individual. These FIBER-FISH confirmed partial AZFb and AZFc deletions were found in 9/246 men with a low sperm count (<10 x 10^6/ml) compared with 0/148 men having a normal sperm count (>40 x 10^6/ml), a difference with distinct statistical significance. It was therefore concluded that despite the frequent gr/gr deletions found also in men from other Y haplogroups, they must be associated with some aspects of male infertility, as is the case for the complete AZFc deletions, although of lower penetrance (Repping et al., 2003).

However, the statistical significance of this conclusion would be lost if the 149th man with normal spermatogenesis in this study would be the first with a gr/gr FIBER-FISH pattern.
with a gr/gr deletion and the D2b haplgroup. Frequent subsequent b2/b4 duplication events would be expected if gr/gr deletions in this Y lineage would have an impact on the D2b men’s fertility. Alternatively, men with a D2b Y chromosome are more likely to have some spermatogenic failure than those with another Y haplogroup.

Although this study suggested that at least some partial AZF deletions have some influence on the men’s reproductive fitness, it is still preliminary since it did not distinguish the g and r amplicon based partial AZFc deletions. Moreover, their influence on the men’s fertility is probably different in different Y lineages and its penetrance with respect to spermatogenic failure is certainly lower than for the complete AZFc deletions induced by b2/b4 recombinations. Whereas complete AZFc deletions were almost always found as ‘de novo’ deletions, i.e., only present in the patient’s Y chromosome (Vogt, 1998; Krausz et al., 2003), in almost instances in which the father of an infertile gr/gr deleted man was available, the father’s Y chromosome was also gr/gr deleted (Repping et al., 2003; Machev et al., 2004).

The first g1/g2 based recombination in AZFf leading to a partial deletion which included the DAZ1/DAZ2 gene doubled was reported by Fernandes et al. (2002). It was found as ‘de novo’ mutation event in five individuals with severe oligozoospermia and was confirmed with a specific DNA blot assay using the association of distinct EcoRV and TaqI fragments from the DYS1 locus to specific DAZ gene copies of the Y chromosome from branch R*. Since the same DAZ1/DAZ2 blot deletion pattern could not be identified in 107 fertile control samples, it was assumed that the g1/g2 deleted Y chromosomes are probably associated with spermatogenic failure. This conclusion has now been confirmed by three similar studies identifying a DAZ1/DAZ2 deletion in two men with incomplete meiotic arrest (Ferras et al., 2004), in ten men with azoospermia or several grades of oligozoospermia (Ferlin et al., 2005), and in four men with severe oligozoospermia (Giachini et al., 2005). Most interestingly, the DAZ1/DAZ2 deletions found in the last study were always associated with deletion of the proximal CDY1 gene copy (Giachini et al., 2005) supporting the g1/g2 recombination mechanism as their putative origin.

Two further studies identified partial AZFc deletions in infertile men and fertile men (Hucklenbroich et al., 2005; de Llanos et al., 2005). Both were, however, based solely on PCR-assays using the STS markers suggested by Repping et al. (2003) for the R1*-Y sequence: sY1258 proximal to b1, sY1161 distal to b1 and b2, sY1197 proximal to t2, sY1191 in u3, sY1291 distal in r2, sY1206 marking the ends of the g2 and g3 palindrome and sY1201 marking the distal border of the complete AZFc deletion interval (Figure 3). Accordingly, and only if the Y chromosome analysed has the same amplicon structure in distal Yq11 as known from the sequence of the R1*-Y chromosome and if no gene conversions have transferred there sites to other palindromic AZF sequence regions, b1/b3 deleted Y chromosomes should be marked by deletion of both sY1161 sites and deletion of sY1197, sY1191, sY1291 but presence of sY1258, sY1206 and sY1201 whereas the gr/gr deleted Y chromosomes are marked by deletion of the sY1291 site with presence of all other STSs. If however, the underlying AZFb/C amplicon structure is different from that in the R*-Y sequence see also Figure 3B and C) deletion of these markers might indicate some other still unknown Y chromosomal rearrangements.
In one STS deletion screen genomic DNA samples of 283 individuals with some spermatogenic failure and of 232 fertile control samples were included (de Llanos et al., 2005). It identified 12 putative gr/gr recombined Y chromosomes by sY1291 deletion in the infertile men population (i.e., 4.24%) and no sY1291 deletion in the Y chromosomes of the fertile men group. The highest frequency of sY1291 (gr/gr?) deleted Y chromosomes was found in the oligozoospermic men group (11/217). This distribution is opposite to that found for complete AZFc deletion with generally a higher incidence in the azoospermic men (Krausz et al., 2003). The authors therefore also proposed that gr/gr deleted Y chromosomes might not be a cause for failure of spermatogenesis but only a distinct risk factor similar like mitochondrial DNA deletions (Kao et al., 1998; Cummins, 2001). However, as stated above, since this study did not confirm the gr/gr association of their sY1291 deletions by appropriate FIBER-FISH or DNA blot experiments or other quantitative gene copy assays and also did not include a Y haplogroup analysis it cannot be excluded that the authors have identified some other Y chromosomal rearrangements in an AZFb/c amplicon structure which is different from that of the R*-Y-sequence and that the sY1291 deletion is not due to some gr/gr based real partial AZF deletion but due to a polymorphic STS deletion as recently found in men with Y haplogroup J (Machev et al., 2004).

Interestingly, a second large (also solely PCR-based) STS screening study on gr/gr and additionally on b1/3 and b2/b3 deleted Y chromosomes, found no significant genotype/phenotype association (Hucklenbroich et al., 2005). Screening 348 men with non-obstructive oligo/azoospermia and 170 men with normal spermatogenesis no impact of the identified sY1291-deleted Y chromosomes on the men’s spermatogenesis profile could be noticed, the sperm count of men with the sY1291-deleted Y variant were all in the normal range. Moreover also b1/b3 and b2/b3 deleted Y chromosomes did not interfere with the fertility status of the men analysed. The authors identified also three novel partial AZFc deletions not fitting with their STS deletion pattern to a b1/b3; b2/b3 or gr/gr recombined AZFb/c amplicon structure of the R1*-Y chromosome. Unfortunately, no FISH or blot analyses were presented to confirm the novel AZFc deletions, nor did a Y haplogroup analysis reveal their possible association to a distinct Y lineage. Results of a Y haplogroup analysis were only reported for the three fertile control samples with a sY1291-deleted Y chromosome and a normal sperm count (Hucklenbroich et al., 2005). They belonged to Y haplogroup R1* (two individuals) and F* (one individual). Both Y branches were most frequent in the population analysed and are known to contain frequently the gr/gr deleted AZFc sequence variant (Repping et al., 2003).

A third large screening study for partial AZFc deletions was based not only on STS assays but also on the quantitative analysis of a number of gene markers for the DAZ and the CDY1 genes including some novel gene copy specific sequence variants. It also included FIBER-FISH experiments to distinguish between the presence of one or more DAZ gene doublets in the men’s lymphocyte nuclei and distinguished the 153 Y haplogroups in the five major branches: Y(xD,E,J,P), DE, P, and J. (Machev et al., 2004) Three different groups were screened for gr/gr deletions: (1) 300 infertile men with sperm counts between 0–131 × 10^6/ml, (2) 210 men of unknown fertility status and (3) 185 fertile men with one or more children. No significant difference was found in the frequency of gr/gr deleted Y chromosomes in the three groups analysed (Machev et al., 2004).

However, considering the control groups with unknown sperm numbers (2 and 3) there might be a bias. If gr/gr deletions are indeed influencing the men’s sperm numbers as originally stated (Repping et al., 2003) the presence of oligozoospermic men must be excluded from these control groups. Only ‘normospermic’ men with normal fertility would be acceptable. A reduced sperm count does not necessarily cause infertility. Therefore, fertile men in the control groups with a gr/gr deletion in AZFc are expected to have a lower sperm count.

Indeed, when comparing the frequency of putative gr/gr-recombined AZFc structures in a group of normozoospermic (n = 189) and of oligo/azoospermic (n = 150) men the frequency of the gr/gr coined partial AZFc deletions were significantly higher in the oligo/azoospermic group (5.3%) than in the normozoospermic controls (0.5%) (P < 0.012) (Giachini et al., 2005). This most recent study also confirmed that there is large heterogeneity in the partial ‘gr/gr’ coined AZF deletions distinguished by a marker of the CDY1 gene and an extreme heterogeneity of the phenotype ranging from azoospermia to normal sperm numbers. Also in the Machev-study (Machev et al., 2004) overall 32 different gr/gr deletions were detected originating from at least 17 independent gr/gr recombination events.

Interestingly, in each Y haplogroup two gr/gr deleted Y chromosomes were identified with an associated b2 ↔ b3 or b3 ↔ b4 inversion (Machev et al., 2004). It can therefore be predicted that the putative inversion events possible in the AZFb/c amplicon structure (Figure 3C) probably occur with a similar frequency as the ampliconic recombination events associated with this Y haplogroup.

**Mapping of AZF deletions in men with a Y(xR) AZF c amplicon structure**

The analysed marker deletion- and/or FIBER-FISH-patterns of some partial AZFc deletions could sometimes only be placed along a continuous DNA segment if some inversions were allowed in the R*-AZFc amplicon structure (Machev et al., 2004) leading to different polarities of the AZFc amplicons (Figure 3C). The b2 ↔ b3 with subsequent gr/gr recombination and associated AZFc deletion results in deletion of all markers of the AZFc-u3 single copy region including the polymorphic 502/C fragment. This b2 ↔ b3 inversion was first proposed to be present in the AZFb/c amplicon structure of men from a pre-N-haplogroup, because a g1/g3 recombination event leading to deletion of the DAZ3/DAZ4 gene doublet was found in all individuals from Y haplogroup N (Fernandes et al., 2004). The g1/g3 deleted AZFc structure was marked by an extensive deletion pattern of STS/printed SNV markers analysing 37 positions along ~3000 kb of the R1*-AZFc amplicon structure and the DAZ3/DAZ4 deletion was confirmed by specific DNA blot analyses. A g1/g3 recombination is not expected in the R1*-AZFc amplicon structure because of the opposite polarity of g1 and g3 in this sequence (Figure 3A).

The same Y-N associated AZF amplicon structure was derived from a variant DAZ-FIBER-FISH pattern found in individuals
with deletion of the sY1191 AZFc-u3 marker (Repping et al., 2004). These authors concluded that not a b2 ↔ b3 inversion but a gr ↔ gr inversion (‘gr-gr’) in the R1*–AZFc amplicon structure with subsequent b2/b3 recombination would result in the Y–N AZFc amplicon structure. However, it must be admitted that both methods used for analysis of the different ampliconic rearrangements (DNA-blot and FIBER-FISH) are inadequate to distinguish both possibilities and that the variance of the FIBER–FISH patterns observed with deletion of the sY1191 marker can also point to different gl/g3 recombination sites in the pre-N*-AZFc amplicon structure.

The best guide to the proposed pre-N-AZFc amplicon structure may be probably provided by the Y chromosomes of the sister clade of the N-branch in the current Y phylogeny, namely haplogroup O (Figure 4). Here a b2 ↔ b3 inversion and not a gr ↔ gr inversion was accounted for the arrangement of the Y–O AZFc amplicon structure (Repping et al., 2004). If this holds true, the Y-N deletion in AZFc would be the result of a gl/g3 recombination based on a pre-N b2 ↔ b3 inversion, rather than of a b2/b3 recombination. The same conclusion was also drawn from the studies of Machev and coworkers and extended to the observation that the same b2 ↔ b3 rearrangement might be not only present in men of Y-N but also in men of Y–F* and Y–I (Machev et al., 2004). This was also found in a study of 1563 individuals by Repping et al. (2004).

It has been already appreciated some years ago that just the sequence area of the u3 AZFc region is particularly vulnerable for genomic rearrangements. At least six independent deletions and four duplications affect a short section of this sequence region which is marked by the polymorphic DYS7-50f2/C fragment (Figure 3A). Together these u3 linked AZFb/c rearrangements are present in ~8% of normal men (Jobling et al., 1996). The u3-marker sY1192 is only ~13.5 kb proximal to the 50f2/C u3-AZFc section (Fernandes et al., 2004). Considering the fact that we probably know now the origin of only three of the proposed six 50f2/C deletions and no origins for the 50f2/C duplication events we must conclude that the molecular base of most of the ten proposed rearrangements in the AZFc-u3 sequence block, including the 50f2/C site, are still unknown.

Based on these earlier analyses and the now known R1*-Y chromosome associated AZFb/c amplicon structure it can therefore be predicted that more rearrangements than described above will exists in at least some of the non-R*-Y haplogroups and that these will lead to more partial AZFb and AZFc deletions not yet identified. Their detection might probably need more sophisticated methods than FIBER-FISH or genomic DNA restriction assays due to the high frequencies of sequence conversions observed in the palindromic organisation of the AZFc amplicon structure (Rozen et al. 2003). To be on the safe side, it seems therefore wise not to restrict any partial AZF deletion analysis to a simple PCR format if one wants to draw some conclusions from the identified AZF-STs deletions to its putative association with the patient’s testicular pathology.

Are some AZFa and AZFb deletions associated with distinct Y haplogroups?

The ‘classical’ AZFa deletions caused after HERV15yg1/y2 recombination events in proximal Yq11 and including both AZFa genes, USP9Y and DBY, are expected to occur with a different frequency in Y chromosomal haplogroups and with and without the L1tr element in the distal HERV15yg2 sequence block (Blanco et al., 2000). The deletion of the 12f2-2 kb sequence of the DYS11 locus has occurred independently in at least two different Y branches. It cuts out precisely the LIPA4 sequence, a truncated LINE element (Kamp et al., 2000). This would result in a substantially longer segment of sequence identity between both HERV15 elements. An increased rate of HERV15yg1/yq2 recombinations resulting subsequently in the classical AZFa deletions would be therefore expected in men with a Y chromosome of the 12f2-2 kb haplogroups (Blanco et al., 2000).

Partial AZFa deletions including only the USP9Y gene (Qureshi et al., 1996; Blagosklonova et al., 2000) or DBY gene (Foresta et al., 2000; van Landuyt et al., 2001) were repeatedly reported but without any analysis of the associated Y haplogroups. Since the complete Yq11 sequence is structured with multiple duplicated sequence blocks shorter than the prominent AZFb/c amplicons but long enough for multiple homologous NAHR based recombinations (Liskay and Stachelek, 1986), it is most likely that these partial AZFa deletions are due to some still unknown NAHR events in the corresponding sequence areas and it would be interesting to learn whether they are associated with one or more distinct Y lineages. This linkage is suggested because the occurrence of partial AZFa deletions seemed to be restricted to distinct populations, they were not found in any large screening surveys for AZF deletions (Vogt, 1998; Simoni, 2001, Krausz et al., 2003).

Considering the polymorphic deletion of the 50f2/E sequence together with the 50f2/C sequence and the RBMY gene copies (Jobling et al., 1996) and presence of the polymorphic LLY22 marker in distal AZFb (Kirsch et al., 1996) a polymorphic structure in the unique distal AZFb sequence, i.e., not overlapping with the AZFc deletion interval, is expected. The variant and polymorphic exon structure of the RBMY gene copies in the same Y region (Prosper et al., 1996, Yen, 1999) and the identification of some unique partial AZFb deletions identified in four infertile Italian men (Ferlin et al. 2003) seem to support this conclusion. The Italian partial AZFb deletion can not be explained by any ampliconic recombination event from the R1*–AZFb/c amplicon structure. An intriguing possibility might therefore be the presence of some chromosomal rearrangements in the distal AZFb sequence associated with the specific Y lineage of these patient’s families. Most interesting, these partial AZFb deletions are probably not a polymorphic neutral deletion event since they were confirmed as a ‘de novo’ mutation in two individuals with distinct testicular pathologies (Ferlin et al., 2003).

Are AZF deletions and duplications two sides of the same NAHR medal?

If NAHR is the molecular mechanism causing subsequently the observed complete and partial AZFa, AZFb and AZFc deletions, it can be predicted that the reciprocal events, AZFa, AZFb and AZFc duplications, will be also generated (Potocki et al., 2000). Molecular duplications are probably rarely pathogenic and harder to detect because they require a quantitative rather than
might be the origin of the seven DAZ earlier in some fertile individuals by FIBER-FISH (Glaeser et al., 2003). Duplications of part of the Yq11 euchromatin forming the dicentric dic-(Yp) or Ynf chromosome are always associated with male infertility (see previous chapter).

Surprisingly, in contrast of this, duplication of the DAZ-AZFc interval and translocation to the proximal part of the Y short arm seems to be compatible with human fertility (Engelen et al., 2003). This Y chromosomal rearrangement reminds to the occurrence of a similar inversion event in the Y chromosome of the Gujarati Muslim Indian population in South Africa (Bernstein et al., 1986). Although there are not yet any molecular experiments supporting this view, it can be assumed that both Y rearrangements are probably based on the similar sequence structure of proximal Yp and the distal AZFb region in Yq11. Looking at the R1*-Y sequence structure (Skaletsky et al., 2003), copies and pseudo-gene copies of the RBMY, TSPY, and PRY genes were mapped to both these Y regions.

Submicroscopically smaller (molecular) AZF duplications can be estimated with different molecular methods: (1) by a semiquantitative analysis of STR markers mapped inside and outside the duplicated AZF region comparing their peak densities in denaturing sequence gels, (2) by a quantitative DNA blot experiment with probes hybridising inside and outside the duplicated AZF region comparing their autoradiographic signal intensities, (3) with FIBER-FISH experiments on the duplicated AZF chromatin domain with Y-specific cosmid clones.

Two duplications of the AZFalpha region mediated by the HERV15 sequence blocks in proximal Yq11 have been revealed during a population survey of 9 Y-STRs mapped in the AZFalpha region (Bosch and Jobling, 2003). Both duplications seemed not to interfere with the men’s fertility status. They might therefore be widespread and present in different Y haplogroups, similar as found earlier for the 50f2/C site in proximal AZFc (Jobling et al., 1996). One of the 50f2/C marked duplications in proximal AZFc might be the origin of the seven DAZ gene copies observed earlier in some fertile individuals by FIBER-FISH (Glaeser et al., 1998). Duplications in AZFc followed after a gr/gr based partial AZFc deletion are the b2/b4 duplications also identified by FIBER-FISH (Repping et al., 2003). It can be concluded that similar polymorphic duplications are present along the whole Yq11 sequence region due to its high amount of repetitive sequence blocks. They might be probably triggered by continuous genomic deletion/amplification mechanisms selected in the different Y haplogroups for balancing their male’s fertility status.

Summary

Obviously we are just beginning to understand the dynamic structure of the human Y chromosome and the putative range of its possible rearrangements. All the observed molecular variations are probably based primarily on its extraordinary repetitive sequence structure. This is especially true for the Yq11 euchromatic sequence region where the AZF locus resides embedded in different polymorphic intervals historically divided in AZFa, AZFb and AZFc (Vogt et al., 1996). Quantitative blot analysis (Kirsch et al., 1996) and then sequence analysis (Skaletsky et al., 2003) has shown that there is a large overlap between the AZFb and AZFc deletion intervals and that most Y genes expressed solely in human testes tissue are deleted with the AZFb deletion (Vogt, 2005).

Considering the variable copy number of these Y genes in Y haplogroup D2b, F(xH,K), I and N and probably more Y haplogroups not yet identified, the question is raised which Y genes in these polymorphic AZFalpha subintervals are really essential for spermatogenesis. That means which gene deletion is really a causative agent for the clinically observed man’s testicular pathology: which gene deletion is neutral (polymorphic) and which AZF, gene is only balancing and shaping the reproductive fitness factor(s) of the male in the different human populations and if deleted is counterbalanced by another beneficial Y or non-Y male fertility factor still unknown (see also Quintana-Murci et al., 2001a; Vogt, 2004).

In this context it is worth to consider also the choice of the Y reference sequence, which was in fact accidentally extracted from a man with an unknown fertility status now known to belong to the R* (Fernandes et al., 2004). If this Y sequence would have been derived from a man in the Y branch N*, our view on the AZFc amplicon structure would have been simpler but we would have also no sequence data from the AZFc-u3 segment. Consequently, the R*-Y sequence now available may also lack sequence regions which are present on the Y chromosome from another Y haplogroup.

In many cases there are alternative recombinatorial routes for the homologous amplicons leading to the same AZFalpha amplicon structure as discussed here for the AZFc structure of Y lineage N*. If derived from the direct polarity pattern of the R*-AZFc amplicon blocks, a b2/b3 inversion followed by a g1/g3 recombination in the pre-N* lineage is most likely (Fernandes et al., 2004; Machev et al., 2004) but another route based on a gr/gr inversion (Repping et al., 2004) or the route based on an inversion of the yellow amplicons (P1.1/P1.2) (Machev et al., 2004) can not be excluded. STS deletion analysis but also FIBER-FISH and blot analysis are probably not suitable to identify and to distinguish these possibilities. Additional analyses of sequences and structural maps from the Y chromosome of other branches –perhaps choosing first the most divergent A haplogroup (The Y Chromosome Consortium (YCC), 2002)– are therefore highly desirable.

Acknowledgements

I like to thank Profs. Ann Chandley (University of Edinburgh), Erika Bühler (University of Basel) and Richard Preiffer (University of Erlangen) for having introduced me in the light complexity of Y chromosomal rearrangements. I am indebted to Chris Tyler-Smith for paving me the way towards a molecular understanding of the complexity of the Y chromosomal haplogroups in different human populations. My students Susana Fernandes, Christine Kamp, Stefan Kirsch, and Mike Koehler and my technicians, Angela Edelmann and Karin Hü llen are thanked for their valuable experimental contributions to the analysis of the distinct Y chromosomal rearrangements in infertile men and fertile men and the associated variabilities of the AZF genes. Mark Jobling is thanked for sharing with me a novel edition of the Y haplogroup pedigree presented in...
AZF deletions and Y haplogroups based on sequence


References


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


Received on February 11, 2005; revised on April 15, 2005; accepted on April 22, 2005