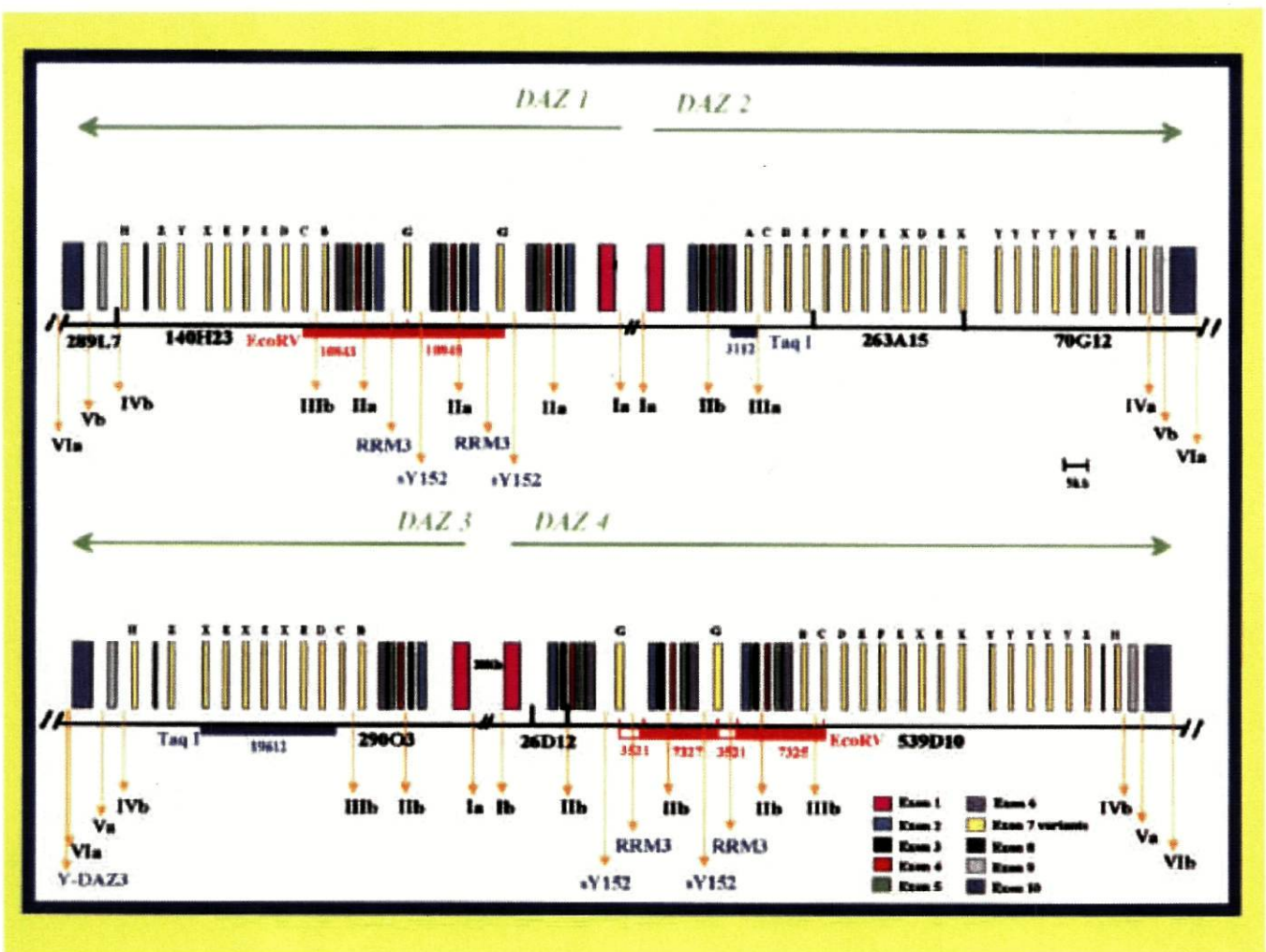
Molecular analysis of the structural and functional complexity of the human *AZFc* male fertility locus in distal Yq11

Análise molecular da complexidade estrutural e funcional do locus da fertilidade masculina localizado na porção distal de Yq11 (*AZFc*)



Susana Fernandes

Porto 2003

Molecular analysis of the structural and functional complexity of the human *AZFc* male fertility locus in distal Yq11

Análise molecular da complexidade estrutural e funcional do locus da fertilidade masculina localizado na porção distal de Yq11 (*AZFc*)

Susana Fernandes

Porto 2003

Dissertação de candidatura ao grau de Doutor
apresentada à Faculdade de Medicina
da Universidade do Porto

Artigo 48º, § 3º - A Faculdade não responde pelas doutrinas expendidas na dissertação. (Regulamento da Faculdade de Medicina do Porto – Decreto-Lei nº 19 337, de 29 de Janeiro de 1931).

Agradecimentos

Ao Professor Doutor Alberto Barros, orientador deste trabalho, o meu sincero agradecimento por todo o estímulo, apoio e orientação científica e pedagógica.

To Prof. Peter Vogt, co-supervisor of this thesis, I would like to thank the enthusiasm towards this work as well as the shared experiences and discussions contributing to my scientific maturing process.

Ao Professor Doutor Mário Sousa, agradeço todo o apoio, toda a confiança depositada no meu trabalho, assim como toda a orientação científica.

Aos co-autores das publicações incluídas nesta dissertação, agradeço a valiosa colaboração.

To the co-authors of the publications included in this work, I would like to thank for all the collaboration.

To Chris Tyler-Smith and his group, I would like to thank for all the work concerning the Y-haplotyping.

To Anne Jauch and her group, I would like to thank all the work concerning the M-FISH studies done in all cell lines used in this work.

À Fundação para a Ciência e Tecnologia e ao Programa PRAXIS XXI, pela Bolsa de Doutoramento atribuída (SFRH/BD/811/2000).

To all my colleagues in the lab in Heidelberg, thank you for all help and useful discussions.

To all my friends (German, Chinese and others) thank you for making me feel like at home! With all of you my time in Heidelberg was more pleasant.

À Mena, pelo incentivo e todo o apoio incondicional.

Aos meus pais e irmãos, por tudo.

Ao abrigo do Decreto-Lei nº 216/92, fazem parte integrante desta dissertação os trabalhos já publicados ou submetidos para publicação:

- Sousa M, Fernandes S, Barros: Prognostic factors for successful testicle spermatid recover. *Molecular and Cellular Endocrinology*, **166**: 37-43, 2000.

- Fernandes S, Huellen K, Gonçalves J, Dukal H, Zeisler J, Rajpert De Meyts E, Skakkebaek NE, Habermann B, Krause W, Sousa M, Barros A, Vogt PH: High frequency of *DAZ*1/*DAZ*2 gene deletions in patients with severe oligozoospermia. *Molecular Human Reproduction* **8 (3)**: 286-298, 2002.

- Vogt PH, Fernandes S: Polymorphic *DAZ* gene family in polymorphic structure of AZFc locus: Artwork or functional for human spermatogenesis? *APMIS* **111**: 115-127, 2003.

- Fernandes S, Paracchini S, Meyer L, Floridia G, Tyler-Smith C, Vogt PH: A large *AZF*c deletion removes *DAZ*3/*DAZ*4 and nearby genes from Y-haplogroup N men. *American Journal of Human Genetics* (in press).

Em cumprimento do disposto no referido Decreto-Lei, declara que participou activamente na recolha e estudo do material incluído em todos os trabalhos, tendo redigido os textos com a colaboração dos outros autores.

A dissertação inclui também resultados ainda não enviados para publicação.

Nota explicativa

A presente dissertação está escrita em inglês na sua quase totalidade, pelo facto do Prof. Peter Vogt ter sido o seu co-orientador.

O genoma é um livro
Tem 23 capítulos, chamados *cromossomas*
Cada capítulo contém vários milhares de histórias, chamadas *genes*
Cada história é feita de parágrafos, chamados *exões*,
que são interrompidos por anúncios chamados *intrões*
Cada parágrafo é feito de palavras, chamadas *codões*
Cada palavra é escrita com letras, chamadas *bases*

(adaptado do livro *Genoma* de Matt Ridley)

Contents

Abbreviations xix

1. Introduction 1

1.1. Spermatogenesis and Y chromosome 3

1.2. Azoospermia factor(s) (*AZF*) 6

1.3. *AZF*c candidate genes 8

1.4. *DAZ* gene family 9

1.5. Y-chromosomal haplogroups and DYS1/*DAZ* locus 11

1.6. LINE transposable elements in the *DAZ* locus 13

1.7. Methylation and acetylation of DNA and chromatin 15

1.8. Aims 17

2. Material 19

2.1. Media for bacterial cultures 21

2.2. Agarose gel electrophoresis 21

2.3. Plasmid, PAC- and BAC-DNA isolation 22

2.4. Genomic DNA isolation 23

2.5. RNA isolation 23

2.6. DNA-blot and radioactive hybridisation 24

2.6.1. Labelling 24

2.6.2. Blotting 24

2.6.3. Hybridisation 24

2.7. Enzymes 25

2.8. cDNA synthesis 25

2.9. Oligonucleotides / Primers 26

2.10. DNA sources 27

2.10.1. Human PAC- and BAC-DNA clones 27

2.10.2. Human genomic DNA 28

2.10.3. Vectors 28

2.10.3.1. pCR 2.1-TOPO 28

2.10.3.2. pBluescript II SK 29

2.10.3.3. pBACe3.6 30

2.10.3.4. pCYPAC2 31

2.10.4. Probes 32

2.11. Cell culture 32
2.12. Photographic material 33
2.13. *In silico* - Sequence analysis 33
2.13.1. Databank 33
2.13.2. Software for sequence analysis 33
2.14. Strains and cell lines 34
2.14.1. Bacterial strains 34
2.14.2. Human cell lines 34
2.15. Sequencing 35
3. Methods 37
3.1. Microbiology techniques 39
3.1.1. Bacterial cultures 39
3.1.2. Bacterial stocks (Glycerol stocks) 39
3.1.3. Screening genomic libraries 39
3.2. Agarose gel electrophoresis 40
3.3. DNA isolation 41
3.3.1. From blood 41
3.3.2. Plasmid preparation 42
3.3.2.1 DNA Mini-Preps 42
3.3.3. BAC-DNA preparation 43
3.4. RNA isolation 44
3.4.1. From blood 45
3.4.2. From cultured cells 46
3.5. Genomic DNA blot experiments 46
3.5.1. Sample preparation 46
3.5.2. Capillary Blot 47
3.5.3. Vacuum-Blot 48
3.6. Radioactive hybridisation 48
3.6.1. Radioactive labelling by Random Priming 49
3.6.2. Probe cleaning 49
3.6.3. DNA-blot hybridisation 49
3.7. Polymerase Chain Reaction (PCR) 50
3.7.1. Standard PCR 50
3.7.2. Standard PCR programme 51

3.7.3. *DAZ* SNV/STS analyses 51
3.8. RT-PCR 52
3.8.1. Reverse Transcription 52
3.8.2. Amplification 53
3.9. TOPO® cloning 53
3.10. Cell culture 54
3.10.1. Culture of human cell lines 54
3.10.2. Trypsinisation 55
3.10.3. Freezing and thawing 55
3.10.4. Methylation studies 55
3.10.4.1. Azacytidine 55
3.10.4.2. Trichostatin 56
3.10.4.3. Bisulfite 57
3.11. DNA sequencing 58
3.12. Photometric quantification 58
4. Results 61
4.1. Mapping and structure of 4 *DAZ* gene copies 63
4.2. Establishment of a set of *DAZ*-SNV/STS 67
4.2.1. SNV/STS in *DAZ*-PACs and *DAZ*-BACs 69
4.2.2. SNV/STS in fertile men (FM) 70
4.2.3. SNV/STS in oligozoospermic patients (OZ) 71
4.3. DNA-Blot analysis of DYS1 locus 73
4.3.1. *Mlu* I blots 77
4.4. AZFc SNV analysis 78
4.5. *DAZ* promoter structure 83
4.6. Transcriptional regulation of the *DAZ* locus by methylation? 85
4.6.1. Cell lines characterisation 86
4.6.2. *DAZ* transcription studies 88
4.6.2.1. RT-PCR on cell lines 88
4.6.3. *DAZ* transcription after demethylation 91
4.7. Bisulfite treatment 94
4.8. Transcriptional analysis of *DAZ*-L1 element 95

5. Discussion **97**

5.1. AZFc structure 99

5.2. *DAZ*1/DAZ2 deletions associated with severe olizoospermia 99

5.3. DAZ3/4 deletions in fertile men 101

5.4. Putative molecular mechanisms for *DAZ* genes deletion 102

5.4.1. *DAZ1/DAZ2* genes deletion 104

5.4.2. *DAZ3/DAZ4* genes deletion 105

5.5. *DAZ* promoter 107

5.6. *DAZ* transcriptional regulation 107

5.7. Conclusion 107

6. References **109**

7. Appendix **121**

8. Papers **131**

Figures

Introduction

Figure 1	Spermatogenesis	3
Figure 2	Interval map of the Y chromosome	4
Figure 3	AZFa, AZFb and AZFc in Yq11	5
Figure 4	Genes on Y chromosome	5
Figure 5	Structure of DAZ gene	10
Figure 6	Y-Haplogroups	12
Figure 7	Taq I blots	13
Figure 8	Line element	14

Material

Figure 9	pCR 2.1 TOPO® vector	29
Figure 10	pBluescript® II vector	30
Figure 11	pBACe3.6 vector	31
Figure 12	pCYPAC2 vector	31

Methods

Figure 13	Capillary blot	47
Figure 14.	Vacuum blot	48

Results

Figure 15	Exon structure of DAZ-BAC contig	65
Figure 16	DAZ exon 7 variants	66
Figure 17	sY152 primers on *DAZ*2 and *DAZ*4	69
Figure 18	DAZ-SNVs analysis	72
Figure 19	Taq I /EcoR V BACs-DNA blots	75
Figure 20	Taq I /EcoR V DNA blots - *DAZ*-del-hap 5	76
Figure 21	Taq I /EcoR V DNA blots - *DAZ*-del-hap 10	77
Figure 22	Mlu I blots (BACs and genomic DNA)	78
Figure 23	AZFc palindromic structure and its genes	79
Figure 24	AZFc palindromic structure and the AZF-SNVs	80
Figure 25	AZFc-SNVs analysis	81

Figure 26	DAZ promoter predictions	84
Figure 27	CpG island in DAZ promoter	85
Figure 28	*DAZ*-SNVs analysis on cell lines	86
Figure 29	AZFc-SNVs analysis on cell lines	87
Figure 30	EcoR V cell lines DNA blot	87
Figure 31	DAZ primers for transcription studies	88
Figure 32	DAZ transcripts on cell lines	89
Figure 33	DAZ transcripts on cell lines II	90
Figure 34	DAZ transcripts on NT2 cell line	92
Figure 35	DAZ transcripts on Ep2102 cell line	92
Figure 36	DAZ transcripts on Du-145 cell line	93
Figure 37	DAZ transcripts on LNCaP cell line	93

Discussion

Figure 38	DAZ genes with 3 RRM	101
Figure 39	Putative recombination mechanisms s in AZFc	103
Figure 40	Putative mechanism for DAZ1/DAZ2	105
Figure 41	Putative mechanism for DAZ3/DAZ4	106

Tables

Material

Table I	Primers	27
Table II	BACs list	27
Table III	PACs list	28
Table IV	Probes list	32

Methods

Table V	Agarose	41
Table VI	*DAZ*-SNVs PCR conditions	52

Results

Table VII	AZFc BAC contig	64
Table VIII	*DAZ*-SNVs/STSs	68
Table IX	*DAZ* del-hap in *DAZ*-PACs and *DAZ*-BACs	69
Table X	*DAZ* deletion haplotypes	70
Table XI	*DAZ* del-hap analyses in FM	71
Table XII	*DAZ* del-hap analyses in OZ	72
Table XIII	*Taq* I and *Eco*R V restriction fragments	74
Table XIV	*AZFc* SNVs/STSs	80
Table XV	*AZFc* SNVs/STSs analyses	82
Table XVI	*AZFc* SNVs/STSs analyses on Y-hg N	83
Table XVII	Promoter prediction programmes	84
Table XVIII	DAZ promoter preditction	84
Table XIX	Karyotypes from cell lines	86
Table XX	Sequencing results from cDNA *DAZ* clones	90

Discussion

Table XXI	Functional AZFc gene copies	103

Abbreviations

A	Adenine
AMP	Ampicillin
AZF	Azoospermic factor
BAC	Bacterial artificial chromosomes
β-ME	β-Mercaptoethanol
bp	Base pair(s)
BPY2	Basic protein Y2
C	Cytosine
cDNA	Complementary deoxyribonucleic acid
CDY1	Chromodomain Y1
CSPG4LY	Chondroitin sulphate proteoglycan 4-like Y
C-terminal	Carboxy-terminal
DAZ	Deleted in azoospermia
DAZ-del-hap	DAZ deletion haplotypes
DAZL1	Deleted in azoospermia like 1
DAZla	DAZ like autosomal
DBY	DEAD box polypeptide, Y chromosome
dCTP	Deoxycytidine triphosphate
dH_2O	Desionised water
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
EDTA	Ethylenediaminotetracetic acid
EtOH	Ethanol
FBS	Fetal bovine serum
FISH	Fluorescence in situ hybridisation
FM	Fertile men
G	Guanine
GOLGA2LY	Golgi antigen 2-like Y
GSP	Gene specific primer
H4	Histone 4
HAT	Histone acetyltransferase

kb	Kilobase (1000 bp)
kDa	Kilodalton
L1	Line 1 element
LB	Luria Bertani
LINE	Long interspersed nuclear elements
LTR	Long terminal repeat
M	Molar
Mb	Megabase
Min	Minute
mM	Milimolar
mRNA	Messenger ribonucleic acid
MOPS	N-Morpholino propanesulfonic acid
MSP	Methylation specific PCR
MSY	Male-specific region of the Y chromosome
Na	Natrium (Sodium)
NEA	Non essential aminoacids
nm	Nanometer
N-terminal	Amino-terminal
OD	Optical densitometry
O.N.	Overnight
ORF	Open reading frame
OZ	Oligozoospermic men
PAC	Phage artificial chromosomes
PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
PEN	Penicillin
PIGD	Preimplantation genetic diagnosis
Pos	Position
RBMY	RNA binding motif in Y chromosome
RNA	Ribonucleic acid
rpm	Rotations per minute
RRM	RNA recognition motif
RT	Reverse transcriptase
R.T.	Room temperature

RT-PCR	Reverse transcriptase-PCR
SCOS	Sertoli cell only syndrome
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SPGY	Spermatogenesis gene locus on Y
SSC	Saline sodium citrate
STR	Short tandem repeats
STREP	Streptomycin
STS	Sequence tagged sites
T	Thymine
TAE	Tris Acetate EDTA buffer
Taq	Thermus aquaticus
Tris	Tris(hidroximethyl)aminometane
TTY	Testis transcript Y
USP9Y	Ubiquitin specific protease 9, Y chromosome
UTR	Untranslated region
UTY	Ubiquitously transcribed tetratricopeptide, Y chr
UV	Ultraviolet

Introduction

1.1. Spermatogenesis and Y chromosome

Human spermatogenesis is a long and complex process requiring about 70 days and involving a succession of different cell types, generated by mitotic and meiotic divisions (Clermont 1966, Dym 1994). In the initial stages, spermatogonia are divided via mitoses, giving rise to primary spermatocytes, which in turn undergo the first meiotic division leading to secondary spermatocytes. Through the second meiosis these cells produce haploid cells (round spermatids), which elongate during the spermiogenesis process (elongated spermatids) and finally are differentiated into mature spermatozoa, by condensation of the chromatin, substitution of histones by protamines and formation of the acrosome and other sperm components (fig. 1). However, our knowledge of the mechanisms regulating spermatogenesis is still poor, and only recently research has focused on the identification of genes specifically involved in its regulation (Hecht 1995).

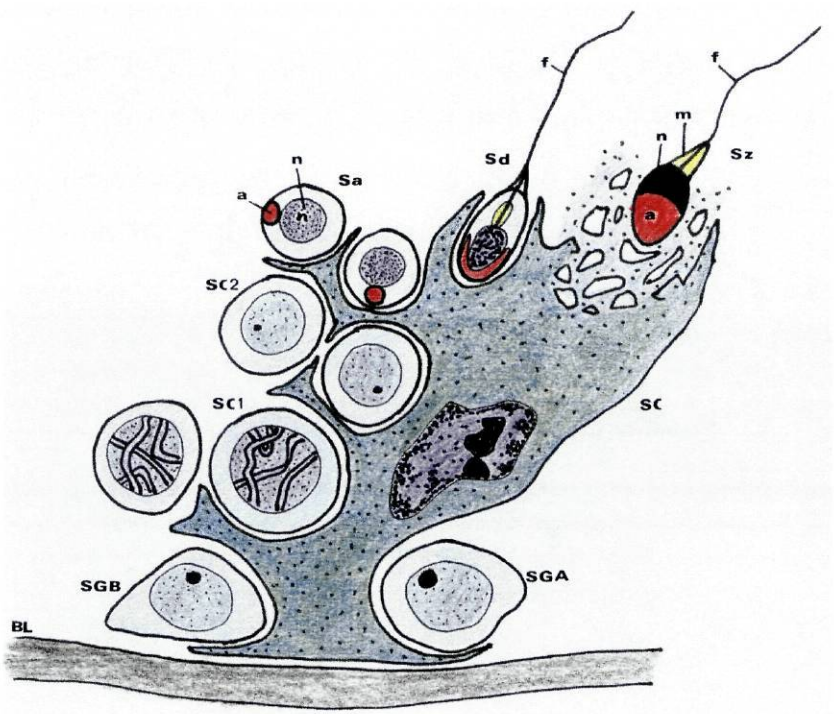


Figure 1 – Schematic representation of human spermatogenesis (Sousa *et al.* 2000). BL - Basal lamina; SC - Sertoli cell; SGA/B - Spermatogonia A and B; SC1 - Primary spermatocytes; SC2 – Secondary spermatocytes; Sa – Round spermatides; Sd – long spermatides; Sz – spermatozoa.

The human Y chromosome shows only one third the size of the X chromosome and constitutes 1% of the haploid human genome with an estimated average size of 60Mb (Morton 1991). Cytogenetically, the human Y chromosome consists of a heterochromatic region and an euchromatic region. The heterochromatic region (Yq12) represents two thirds of the long arm (Yq), is mainly composed of repetitive satellite DNA - DYZ1 and DYZ2 - having, respectively, about 5000 and 2000 sequence copies. The size of this region varies extensively in different males (Unnerus *et al.* 1967). The size of the euchromatic region including the short arm (Yp11), centromere and proximal long arm (Yq11) is quite constant among healthy men. With the exception of the small pseudoautosomal regions (PARs) at the top of both arms (Burgoyne 1998), genetic linkage analysis for the Y chromosome is impossible because of the lack of meiotic recombination. Consequently, the identification of Y-linked genes depends on physical mapping of deletions occurring naturally. To date, two extensive deletion maps of the Y chromosome have been constructed (Vollrath *et al.* 1992, Vogt *et al.* 1996). Vollrath established the first STS interval map of the human Y chromosome. Using 110 STS loci, he was able to subdivide Yq11 into 23 intervals (5A-5Q and 6A-6F) (Vollrath *et al.* 1992). This was later used for gene mapping (fig. 2). The Vogt STS-map (fig. 3) divides Yq11 into 25 intervals (D1-D25), and is based on PCR analysis and DNA blot experiments (Vogt *et al.* 1996).

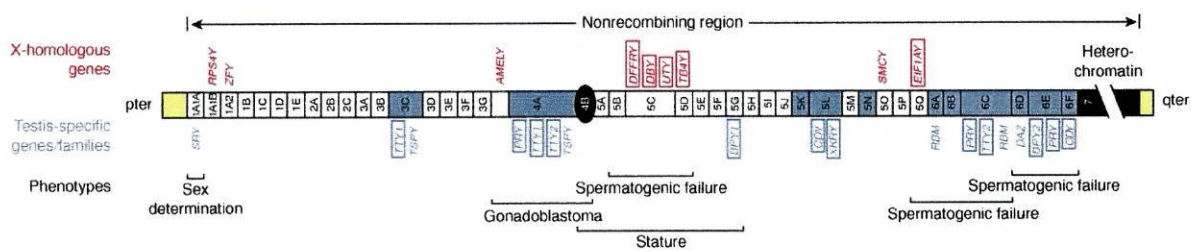


Figure 2 – Interval map of the human Y chromosome according to Lahn & Page (1997).

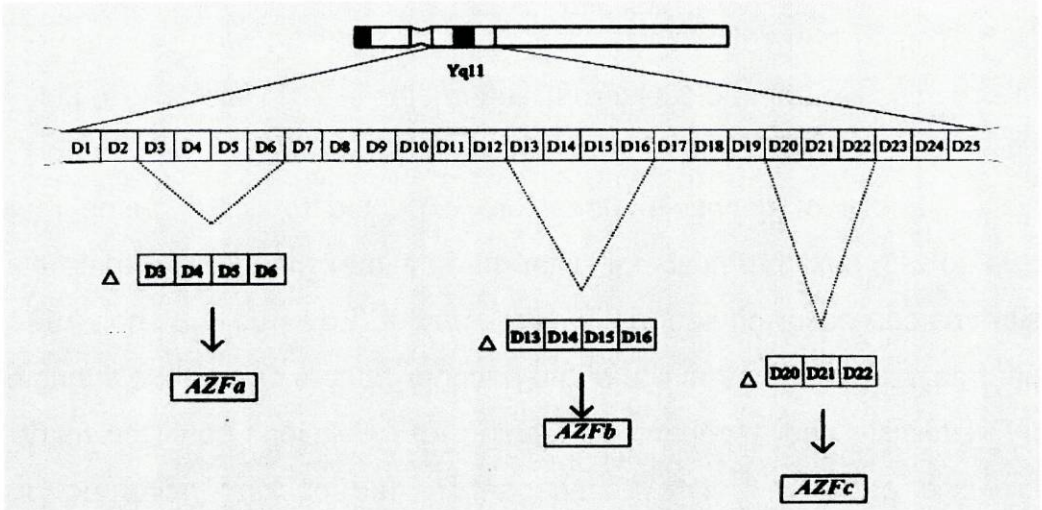


Figure 3 - *AZFa*, *AZFb* and *AZFc* in molecular interval map in Yq11 according to Vogt *et al.* (1996).

The Y chromosome has long been considered as a wasteland devoid of genetic material. This idea was supported by data showing that the euchromatic, non-recombining region of this chromosome is full of Y-specific repeats (Vollrath *et al.* 1992). These sequences were assumed to be functionally inactive (Charlesworth 1991) but, from recent studies, it emerges that the human Y chromosome is far from being a land of nearly dead or degenerate genes but contains multiple genes that are ubiquitously expressed and those with a testis-specific expression (fig. 4).

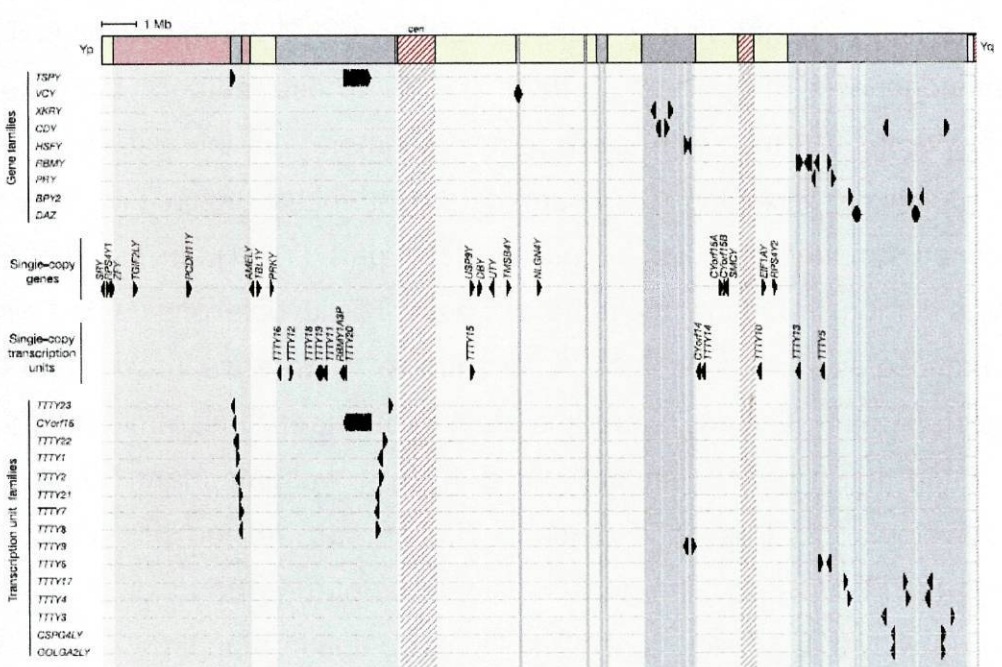


Figure 4 – Genes present on Y chromosome according to Skaletsky *et al.* (2003). Solid triangles denote intact genes, assembled in 4 groups: nine families of protein-coding genes, single-copy genes, single-copy transcription units and families of transcription units.

At the beginning of my thesis work, in 2000, little was known about the origins and mechanisms of the molecular rearrangements in Yq11 resulting in deletions and loss of Y-specific genetic material. Y chromosomal rearrangements and possibly the loss of genetic material are expected to occur during meiosis associated with sex chromosomal sister chromatid exchanges. Further deletion possibilities can occur on spermatid development (Edwards & Bishop 1997), in the spermatozoa of the germ line of the patients' fathers or *de novo* during early embryogenesis. Other mechanisms leading to Y deletions could be related to the specific patterns of DNA condensation and packing occurring during spermiogenesis.

Recently, Y chromosome became completely sequenced and the male-specific region of the Y chromosome (MSY) was defined as a mosaic of heterochromatic sequences and three classes of euchromatic sequences: X-transposed, X-degenerate and ampliconic. The ampliconic class represents about 30% of the euchromatic region of MSY (Skaletsky *et al.* 2003).

1.2. Azoospermia Factor(s) (AZF)

The observation that loss of most distal euchromatic region of the long arm of the Y chromosome is associated with severe oligozoospermia or azoospermia was the first evidence that genetic defects on the Y chromosome can affect spermatogenesis (Tiepolo & Zuffardi 1976). These authors examined the karyotype of 1170 men with azoospermia; six men from this patient group had large deletions, including the entire heterochromatic region (Yq12), and an undefined amount of adjacent euchromatic part (Yq11). In two cases, the authors demonstrated that the fathers of patients with deletions carried a normal Y chromosome, indicating that these mutations were *de novo* events. This suggested that deletions were the cause of the azoospermia and it was postulated that a genetic factor located in Yq11 was important for male germ cell development. This gene or gene cluster was defined as "AZoospermia Factor" (*AZF*). However, the genetic complexity of the *AZF* locus could only be revealed with the development of molecular deletion mapping. These studies enabled the detection of interstitial sub-microscopic deletions not visible at the cytogenetic level and detectable only by Sequence Tagged Sites (STS)-PCR or

Southern hybridisation. Such deletions were called microdeletions. Molecular mapping analysis on patients with microdeletions confirmed that *AZF* resides in Yq11 and further assigned this *AZF* locus to interval 6 at band Yq11.23 (Vogt *et al.* 1992). Heterogeneity in the histology of testis sections from patients with different deletions in Yq11 complicated the original hypothesis of a single locus for spermatogenesis in Yq, suggesting the presence of more than one *AZF* locus in Yq11 (Vogt *et al.* 1993).

Indeed the observed microdeletions are associated to a wide range of histological profiles, from Sertoli cell-only syndrome (SCOS) to spermatogenic arrest (SGA) and severe hypospermatogenesis: deletions in proximal Yq11 (D3-D6) result in type I SCOS, with no germ cells being visible in all tubules of testis sections, whereas microdeletion in middle Yq11 (D13-D16) revealed spermatogenic arrest. In this case, populations of spermatogonia and spermatocytes were normal in all testis tubules. No post-meiotic germ cells could be detected, indicating that the disruption of spermatogenesis occurred before or during meiosis at the spermatocyte stage. Finally, microdeletions in distal Yq11 (D20-D22) are associated with a variable phenotype, ranging from type II SCOS (absence of germ cells in most testis tubules) to hypospermatogenesis (presence of all germ cell types, albeit in reduced numbers). Based on these results, Vogt *et al.* (1996) proposed the presence of three spermatogenesis loci in Yq11 designated as *AZFa*, *AZFb* and *AZFc*, from proximal to distal Yq (fig. 3).

According to Vogt *et al.* (1997), *AZFa*, located on D3-D6, corresponds to the deletion interval 5C of the Vollrath map, *AZFb* spans the interval D13-D16 and corresponds to interval 5O-6B of the Vollrath map and *AZFc*, located on D20-D22, corresponds to interval 6C-6E of the Vollrath map. Several genes located in *AZF* regions are expressed in the testis and could therefore be considered as "*AZF* candidate genes".

The first *AZF* candidate gene - *RBMY* (RNA binding motif in Y chromosome) - was isolated in 1993 (Ma *et al.* 1993) from a region later shown to correspond to *AZFb*. In 1995, a second *AZF* candidate gene - *DAZ* (Deleted in Azoospermia) - was described as a single copy gene identified on the *AZFc* region (Reijo *et al.* 1995). However, a homologous gene designed as *SPGY* (Spermatogenesis gene locus on the Y) and mapped to the same Y region was described as a

multicopy gene (Maiwald *et al.* 1996). Both genes encode testis specific RNA binding proteins. *DAZ* contains seven tandem repeats of a 72 nucleotide unit, known as exon 7 of *DAZ* (Saxena *et al.* 1996); *SPGY* contains at least 12 of these repeats. Further studies revealed that these 2 genes belong to the same family and because the name *DAZ* was published first, *DAZ* was renamed *DAZ*1 and *SPGY* became *DAZ*2 (Vogt *et al.* 1997). Only recently the structure of *AZFa* and its gene content was described, having at least 3 genes - *USP9Y* (Ubiquitin specific protease 9 on Y chromosome), DBY (DEAD box polypeptide on Y chromosome) and UTY (Ubiquitously transcribed tetratricopeptide on Y chromosome) - considered to be *AZFa* candidates (Brown *et al.* 1998, Mazeyrat *et al.* 1998, Sargent *et al.* 1999, Sun *et al.* 1999).

1.3. *AZFc* candidate genes

Microdeletions involve frequently the *AZF*c region, with such deletions being responsible for 10-15% of idiopathic azoospermia and 5-10% of severe oligozoospermia.

At the beginning of this work (2000), three different Y gene families were mapped to the AZFc locus, expressing in human testis and being candidates for *AZFc*: the basic protein Y2 (*BPY*2), the chromodomain Y1 (*CDY*1) and the deleted in azoospermia (*DAZ*) gene family (Lahn & Page 1997, Vogt 1998, fig. 2).

The *BPY*2 gene codes for a 106-residue basic protein (Lahn & Page 1997), spans 21 kb of genomic sequence and consist of 9 exons. A cDNA sequence (GeneBank AF000980) was identified as testis specific and positive clones for *BPY*2 were sequenced and compared to this cDNA in order to detect intron/exon junctions (Stuppia *et al.* 2000).

The human chromodomain Y (*CDY*) is a gene family with, at least, three members named *CDY*1 major, *CDY*1 minor and *CDY*2. The *CDY* family encodes a protein containing a chromatin-binding domain and a potential CoA-substrate catalytic domain (Lahn & Page 1999). Most of the putative protein encoded by the *CDY*1 minor transcript is identical to that encoded by the major transcript except that its carboxyl terminus is divergent. The *CDY*1 genes are mapped to interval 6F and the *CDY*2 to interval 5L of the Y chromosome (fig. 2).

Although it is not yet known the exact number of *CDY* copies present in the human Y chromosome, it has been demonstrated that *CDY* is specifically expressed in adult testis. It was speculated that *CDY* might modify chromatin during spermatogenesis with its deletion contributing to spermatogenic defects observed in *AZF*c-deleted men (Ma *et al.* 2000). Recently it was shown that proteins of these genes have *in vitro* HAT (histone acetyltransferase) activity. The expression pattern and subcellular localisation of the *CDY* proteins were consistent with a role in mediating histone H4 hyperacetylation during spermatid maturation (Lahn *et al.* 2002).

1.4. *DAZ* gene family

DAZ (Deleted in Azoospermia) locus was the first described in this region. *DAZ* is found only on the Y chromosome in humans and Old World monkeys (Saxena *et al.* 1996, Shan *et al.* 1996, Gromoll *et al.* 1999). In all other mammals it is represented as a single copy gene, autosomally located (Reijo *et al.* 1996, Cooke *et al.* 1996). DAZ was acquired by the Y chromosome from this autosomal homologue *DAZL*1.

In model organisms, genetic studies have demonstrated that autosomal *DAZ* homologues play essential roles in germ cell development (Eberhart *et al.* 1996). In *Drosophila*, males mutant for the *DAZ* homologue *boule* are infertile with germ cell arrest at the G2/M transition into meiosis I (Burgoyne 1996, Eberhart *et al.* 1996). The mouse homologue of *boule* is located on chromosome 17 (Cooke *et al.* 1996) and it was designated *DAZla* (*Daz* like autosomal). In mice, the disruption of the *Dazh* gene leads to germ cell loss before birth, rendering both males and females infertile (Ruggiu *et al.* 1997), and a human *DAZ* transgene can partially rescue the sterile phenotype of a knockout mouse for the homologous gene *Dazl* (Slee *et al.* 1999).

DAZ encodes a testis-specific protein with a RNA recognition motif (RRM) at the N-terminus and an auxiliary C-terminal domain containing different number of repeats of a 24 amino acid unit, the so called *DAZ* repeats (Reijo *et al.* 1995, Saxena *et al.* 1996). The *DAZ* transcription unit seems to contain different number and composition of exons and to span about 42 kb. Exon 1 contains the

initiator codon, exons 2 to 5 encode the RNA-binding domain and each exon 7 variants encodes a *DAZ* repeat (Saxena *et al.* 1996, fig. 5).

DAZ transcription is restricted to the germ line (Reijo *et al.* 1995), occurring in spermatocytes and early spermatids (Vogt 1996), with the presence of several *DAZ* transcripts belonging to the same individual and differing in both the copy number and the order of the *DAZ* repeats suggesting that more than one *DAZ* gene is transcribed (Yen *et al.* 1997).

The exact number of the *DAZ* genes in the family is still not clear due to the extremely high similarity (>99%) among the members of the *DAZ* gene family (Saxena *et al.* 1996). However, at least three copies have been indicated by Southern blotting or restriction mapping (Yen *et al.* 1997, 1998) and seven copies, representing either genes or pseudogenes, were proposed by fibre-FISH techniques (Gläser *et al.* 1998). Therefore, the main problem in understanding the biological function of *DAZ* and the genotype-phenotype relation arose from the multicopy nature of this gene.

DAZ genes are deleted in about 15% of non-obstructive azoospermic or severely oligospermic males (13%) and are expressed exclusively in germ cells (Menke *et al.* 1997, Habermann *et al.* 1998).

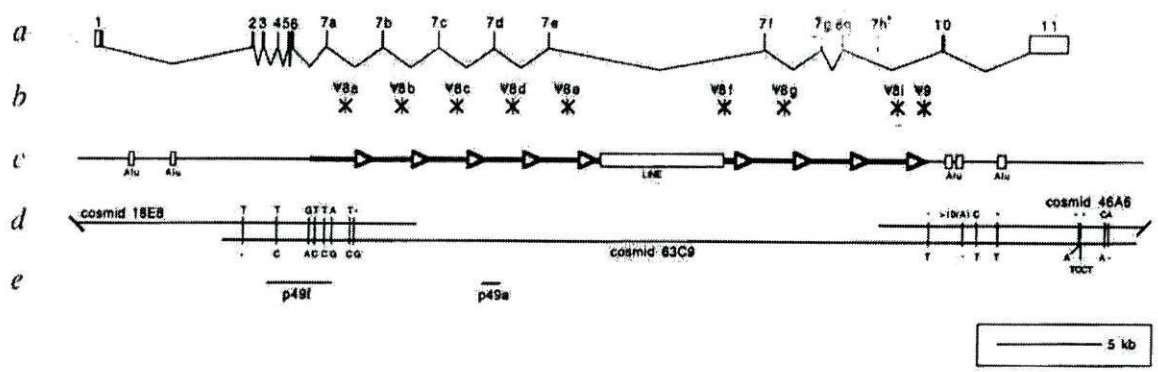


Figure 5 – First exon-intron structure for a DAZ gene copy. a-*DAZ* transcription unit with the exons numbered above it; b-Pseudoexons within *DAZ*; c-Sequence backbone with each arrow showing a 2.4 kb repeat unit and a 6 kb LINE element in between 2 tandem repeat units; d-Three cosmids from which sequence was derived. The nucleotide differences are listed, giving already an indication for the existence of more than one gene copy; e-Localisation of p49f and p49a (Reproduced from Saxena *et al.* 1996).

Deletions of *DAZ* genes in infertile patients are generally screened by PCR on genomic DNA isolated from peripheral leukocytes, using STS markers that cannot distinguish the different copies (Reijo *et al.* 1995, 1996). Therefore, only deletions removing the whole *DAZ* gene cluster can be detected. Intragenic deletions, point mutations or partial *DAZ* deletions (involving only some *DAZ* copies) cannot be determined by this method and there is the possibility that the frequency of *DAZ* gene abnormalities in fertile men might be underestimated.
Since *DAZ* genes seem to be distinguished only by small nucleotide variations, the only way to develop a gene copy specific marker system would be the establishment of the so called *DAZ* deletion haplogroups.

1.5. Y-chromosomal haplogroups and DYS1/*DAZ* locus

An increasing number of informative Y polymorphic markers, i.e. base-pair substitutions, insertion/deletion events, microsatellites, minisatellites and repeat variabilities, have been now isolated (Underhill *et al.* 1997, 2000, Rosser *et al.* 2000). Both biallelic variants and insertion/deletion polymorphisms were recognised as unique molecular events in human evolution and can therefore be used to define deep-rooted phylogenetically related variant pattern called haplogroups (fig. 6).
Microsatellites or short tandem repeats (STR) are more variable and define the variation within distinct haplogroups, allowing the establishment of Y-chromosomal pedigrees developed during human evolution (Jobling *et al.* 1998).

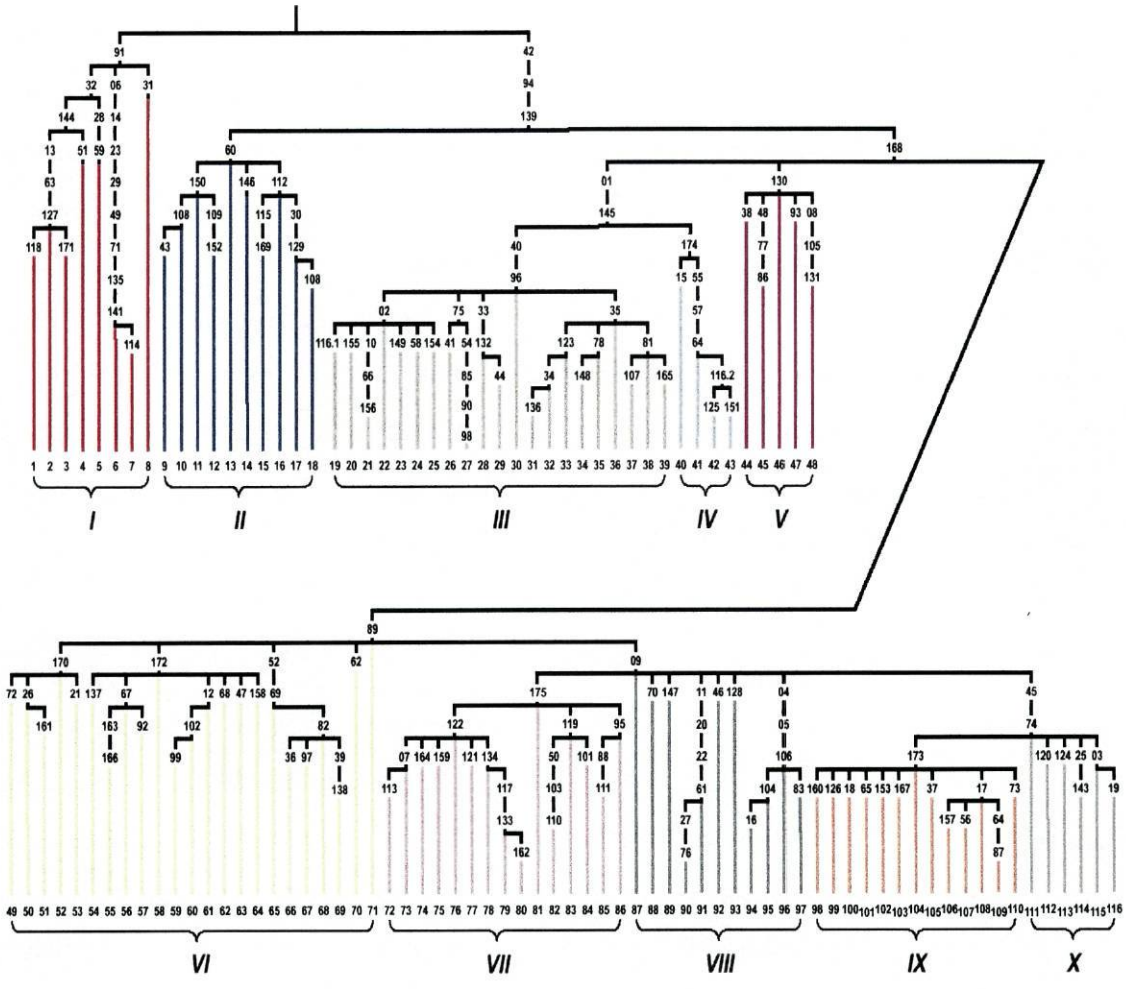


Figure 6 – Y-chromosomal haplogroups according to Underhill *et al.* (2000). The 131 numbered compound haplotypes were constructed from 218 mutations, indicated on segments. Haplotypes are assorted into 10 groups (I-X).

One of the first Y chromosome polymorphic systems used for population studies was the DYS1 locus detected by probes 49a/49f and Southern blots (Ngo *et al.* 1986). These probes recognised up to 18 fragments in *Taq*I digested genomic DNA, named A to R, 8 of which could be present, absent or present in different sizes (Ngo *et al.* 1986, fig. 7). Nowadays we know that the p49a and p49f probes are 6kb distant from each other and are located in the *DAZ* locus (Saxena *et al.* 2000). Therefore the p49a,f haplotypes define individual Y chromosomes (fig. 7).

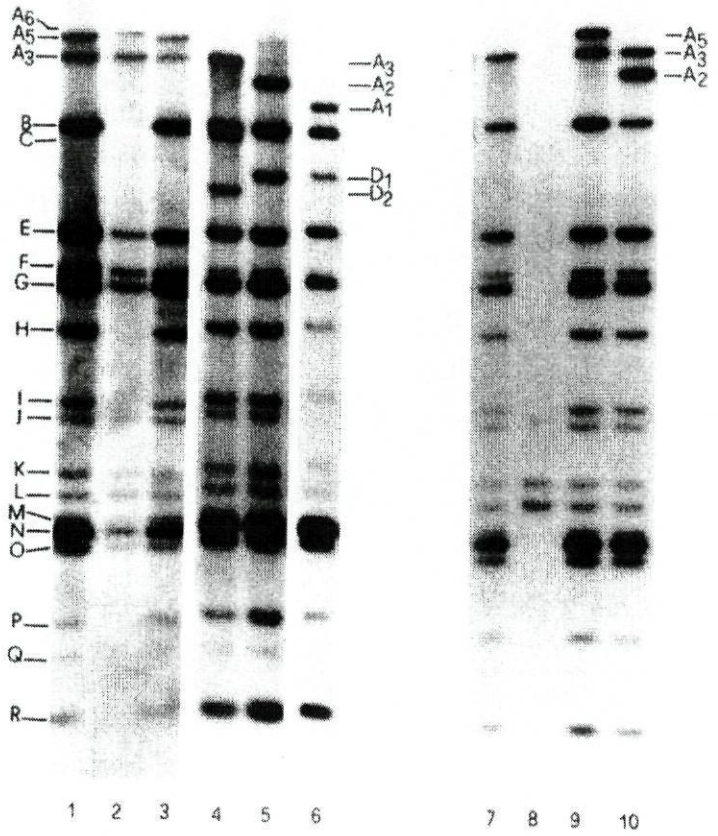


Figure 7 - Electrophoretic patterns of human male genomic DNA digested with *Taq*I and hybridised with p49a/f, according to Ngo *et al.* (1986). The different DNA fragment sizes are named A to R. The variations shown by the B, C, F and I fragments consist in their presence or absence, whereas A and D fragments also show differences in size and are named respectively A1 to A6 from the smallest to the bigger size and D1 bigger than D2. Lanes 1-7, 9, 10 show patterns from different haplotypes. Lane 8 presents a female pattern with only fragments K and L being present.

1.6. LINE transposable elements in the *DAZ* locus

Transposable elements are segments of DNA with the ability to move from one place to another in the genome. They fall into two main classes: retrotransposons, which move by reverse transcription of an RNA intermediate, and transposons, which move by a process of excision and insertion (Wright & Finnegan 2001).

All transposable elements are repeated within the genomes in which they are found, since transposition usually leads to an increase in copy number, either because its mechanism is intrinsically replicative or because of secondary events such as the repair of double-stranded breaks generated during transposition (Wright & Finnegan 2001).

Retrotransposons can be further divided in retrotransposons with long terminal repeats (LTRs) at their ends, equivalent to those of retrovirus, and retrotransposons with no repeats also called LINE elements (Wright & Finnegan 2001).

LINE-1 elements (L1s) are found in all mammalian genomes (Fanning & Singer 1987) and can be distinguished by several structural features. Typically, full-length elements are between 5 and 7 kb long, have no terminal repeats, possess one or two long open reading frames (ORFs) on the strand that terminates in a 3' A-rich segment and are surrounded by variable-length target site duplications (Swergold 1990).

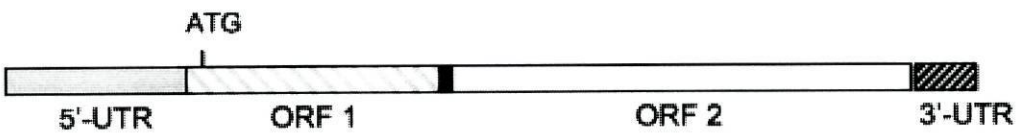


Figure 8 – Line structure containing two ORFs (ORF1 and ORF2) flanked by two untranslated region (5' UTR and 3' UTR), according to Swergold (1990). The dark bar between ORF1 and ORF2 represents the spacer DNA that separates the two ORFs. The start codon (ATG) is shown within ORF1.

The human L1 consensus sequence (Scott 1987) has a 5' untranslated region (UTR) with an internal promotor, 2 open reading frames (ORF1 and ORF2) and a 3' UTR which terminates in a poly A tail (Scott 1987). ORF1 encodes a 40kDa nucleic acid binding protein, while ORF2 encodes a protein with endonuclease (Feng *et al.* 1996) and reverse transcriptase activities (Mathias *et al.* 1991)

The most commonly proposed mechanism of L1 transposition involves (i) synthesis of full-length, polydenylated transcripts, (ii) reverse transcription of the RNA by an L1-encoded enzyme and (iii) insertion into staggered chromosomal breaks (Swergold 1990).

The *DAZ* gene locus contains a full-length L1 element spanning 6.1 kb, between exon 7X and 7Y of each *DAZ* gene copy (fig. 5).

1.7. Methylation and acetylation of DNA and chromatin

DNA methylation is one of the features that organises and controls the expression of genetic information, also having an important role in DNA repair and genome stability (Jost 1993). Methylation of cytosines in certain regions of genes can inhibit gene expression and artificial demethylation of genes has been shown to result in gene reactivation (Chen *et al.* 1976). Differentially methylated regions are key elements in the tissue-specific transcriptional regulation of genes. Methylation was proposed to contribute to the reduction of background expression and to result into the division of complex genomes into transcriptional active and non-transcriptional active regions (Grunau *et al.* 2000).

Mammalian cells possess the capacity to epigenetically modify their genomes via the covalent addition of a methyl group to the 5-position of the cytosine ring within the context of a CpG dinucleotide (Bird 1992). Certain regions of the genome, often clustered at the 5' ends of genes, possess a high CpG frequency (70%) and have been termed CpG islands. DNA methylation can cause an increase in mutation rates and heritage of methylation silenced genes are driven through changes in promoters that are associated with CpG islands (Rideout *et al.* 1990).

In vertebrates, 60-90% of all CpG are methylated and many of the remaining non-methylated CpGs (about 15% of all CpGs in human DNA) are found in CpG islands, which usually include functional promoters (Ng & Bird 1999).

Whereas a number of studies gave evidence for a correlation between the expression state and methylation strength, others showed that demethylation does not necessarily result in transcriptional activation and that the promoter region of inactive genes can be unmethylated. Thus, (de)methylation is apparently no general mechanism to determine tissue-specific gene expression.

Early experiments demonstrated that artificially (5-azacytidine) methylated DNA can adopt a distinct chromatin structure upon integration into the genome (Keshet *et al.* 1986). This conformation renders the DNA refractory to nuclease action. On the other hand, unmethylated CpG islands, associated to promoters of many genes, possess a nuclease-sensitive chromatin structure differing from the bulk of the methylated genome (Tazi & Bird 1990).

Besides DNA methylation, also chromatin methylation and acetylation is known nowadays to remodel the chromatin for silencing a transcriptional unit (Wolffe & Pruss 1996).

In mammals, chromatin remodelling occurs during spermatogenesis, whereby histones are displaced from chromatin, first by transition proteins and later by protamines (Wouters-Tyrou *et al.* 1998). With protamines, DNA in mature spermatozoa is packaged into an extremely condensed, functionally inert configuration (Balhorn 1982). Several evidences led to the view that chromatin remodelling during spermatogenesis is facilitated by hyperacetylation of histone H4. In numerous vertebrates, H4 hyperacetylation correlates to the occurrence of the histone-to-protamine transition in spermatogenesis. Extensive H4 hyperacetylation occurs in the testis of species where histone are replaced by protamines during spermiogenesis, but not in species that completely retain somatic-type histones in mature spermatozoa (Kennedy & Davies 1980). The postulated role of histone hyperacetylation in facilitating histone displacement is consistent with the general perception that histone acetylation results in a more open chromation structure and hence easier access by regulatory proteins. Indeed, it has been observed in the trout that H4 hyperacetylation during spermiogenesis results in a highly relaxed chromatin configuration (Christensen *et al.* 1984) .

The acetylation of N-terminal lysine residues in core histones has been implicated in three distinct cellular processes. The first is the deposition of free histones onto newly synthesised DNA (Parthun *et al.* 1996), the second is the regulation of gene expression and the third is the displacement of histones by transition proteins and protamines during vertebrate spermatogenesis (Meistrich *et al.* 1992).

1.8. Aims of this research

1. Establishment of a contig map of BAC and PAC clones isolated from the *AZFc* region and analysis of their repetitive substructure.
2. Mapping of exon-structures of the AZFc gene families using BAC sequence contig (after sequences became available in GenBank).
3. Design of molecular methods able to distinguish the different copies of the AZFc gene families. Analyses of their deletions in men with normal fertility (AZFc gene number polymorphism).
4. Identifying of the copies of the *DAZ* gene family with a putative function in human spermatogenesis.
5. Mapping of conserved transposable elements in the *DAZ* gene family.
6. Mapping of DAZ gene promoter domain in functional DAZ gene copies.
7. DAZ gene expression analyses in testicular tumour cell lines, as a molecular model for functional DAZ promoter analyses.
8. Origins and mechanisms of the molecular rearrangements in the *AZFc* region resulting in deletions and loss of variable Y-specific genetic material.

Material

2.1. Media for bacterial cultures

LB-Medium	10 g NaCl 10 g Bacto Tryptone (Becton Dickinson) 5 g Yeast Extract (GIBCO BRL) ad 1 l dH_2O, autoclave
LB-Agar for Agar plates	Add 15 g/l Agar to LB-Medium, (Select Agar, Becton Dickinson)
Ampicillin	solution in dH_2O, sterile filter, aliquots at –20°C stock concentration: 50 mg/ml work concentration: 100 µg/ml
Chloramphenicol	solution in 100% Ethanol, sterile filter, aliquots at –20°C stock concentration: 12.5 mg/ml work concentration: 12.5 µg/ml
X-Gal	solution in 100% Ethanol aliquots at –20°C stock concentration: 50 mg/ml work concentration: 40 µg/ml
IPTG	solution in dH_2O, sterile filter, aliquots at –20°C stock concentration: 1 M Work concentration: 0.5mM

2.2. Agarose gel electrophoresis

Agarose	Seakem® LE-Agarose (FMC) LMP Agarose (FMC)
10 x TAE (DNA gel)	40 mM Tris-Acetate pH 7.8 1 mM EDTA
10 x MOPS (RNA gel)	20 mM MOPS 5 mM Na-acetate 1 mM EDTA pH 7.0
Loading buffer Orange G	20 ml Glycerin (87%) 1 ml 1 M Tris, pH 7.5 0.1 g Orange G (AppliChem) up to 100 ml with dH_2O

Loading buffer Bromophenol blue	0.5% Bromophenol blue 50% Glycerin 50 mM EDTA up to 100 ml with dH_2O
RNA dye loading buffer	Novagen # 70606-3
Ethidium bromide	10 mg/ml Sigma

Molecular Weight Markers (MBI Fermentas)

Gene Ruler™ 100bp DNA ladder plus	3000 / 2000 / 1500 / 1200 / 1031 / 900 / 800 / 700 / 600 / 500 / 400 / 300 / 200 / 100 bp
Gene Ruler™ 1kb DNA ladder	10000 / 8000 / 6000 / 5000 / 4000 / 3500 / 3000 / 2500 / 2000 / 1500 / 1000 / 750 / 500 / 250 bp
Lambda DNA / *Hind* III Marker	23130 / 9416 / 6557 / 4361 / 2322 / 2027 / 564 bp

2.3. Plasmid, PAC- and BAC-DNA isolation (Alkaline lysis)

Solutions from Qiagen® Plasmid Mini-prep kit (# 12125) and from Qiagen® Large-construct kit (# 12462):

Resuspension buffer (P1)	50mM Tris-HCl, pH 8.0 10 mM EDTA 100 µg/ml RNase A Store at 2-8 °C
Lysis buffer (P2)	200 mM NaOH, 1% SDS
Neutralisation buffer (P3)	3 M Na-Acetate, pH 5.5
Equilibration buffer (QBT)	750 mM NaCl 50 mM MOPS, pH 7.0 15% Isopropanol 0.15% Triton® X-100
Wash buffer (QC)	1 M NaCl 50 mM MOPS, pH 7.0 15% Isopropanol
Elution buffer(QF)	1.25 M NaCl 50 mM Tris-HCl, pH 8.5 15% Isopropanol

Qiagen® Large-Construct Kit:

Reaction buffer (EX)	50 mM Tris-HCl, pH 8.5 10 mM $MgCl_2$
ATP-dependent Exonuclease	lyophilised, resuspended in 225 µl of Exonuclease-buffer
Exonuclease-buffer	20 mM KCl 20 mM Na-Acetate, pH 7.5
ATP	Stock-solution: 100 mM in dH_2O Store at –20°C in 300 µl Aliquots

2.4. Genomic DNA isolation

Lysis buffer	155 mM NH_4Cl 10 mM $KHCO_3$ 0.1 mM EDTA pH 7.4
SDS	20% in dH_2O
SE-buffer	75 mM NaCl 25 mM EDTA pH 8.0
Proteinase K	10 mg/ml Proteinase K in SE-buffer
TE-buffer (10:1)	10 mM Tris-HCl, pH 8.0 1 mM EDTA

All solutions were sterile filtered (0.2 µm Millipore) or autoclaved.

QIAquick Gel Extraction kit	Qiagen® # 28704
QIAquick PCR purification kit	Qiagen® # 28106

2.5. RNA isolation

QIAamp® RNA blood Mini kit	Qiagen® # 52304
Rnase-Free DNase set	Qiagen® # 79254

2.6. DNA-blot and radioactive hybridisation

2.6.1. Labelling

dATG-Mix	33 mM dATP (Hybaid) 33 mM dTTP (Hybaid) 33 mM cGTP (Hybaid)
$\propto$ ^{32}P-dCTP	SA 3000 Ci/mMol (Amersham Pharmacia Biotech)
Hexanucleotides mixture	2.5 mg/µl (Amersham Pharmacia Biotech)
Probe Cleaning	Sephadex®-G50 (NICK™ Column, Amersham Pharmacia Biotech) #52-2076-00

2.6.2. Blotting

Depurination solution	0.25 M HCl
Denaturation solution	0.5 M NaOH / 1.5 M NaCl
Neutralisation solution	1 M Tris (pH 8) / 1.5 M NaCl
Nylon membrane (Hybond-N+)	Amersham Pharmacia Biotech # RPN 203B

2.6.3. Hybridisation

Church-buffer	0.5 M Na_2HPO_4, 1 mM EDTA, pH 7.2 with H_3PO_4 7% SDS
Salmon Sperm DNA	10 mg/ml Sigma # D1626
20% SDS (Sodium Dodecyl Sulfate)	20 g in 100 ml dH_2O
20 x SSC	3 M NaCl / 0.3 M Na-citrate, pH 7.0
Wash buffer I	2 x SSC 0.1% SDS
Wash buffer II	1 x SSC 0.1% SDS
Strip-solution I (mild strip)	1 mM Tris-HCl, pH 8.0 1 mM EDTA, pH 8.0 0.5 % SDS
Strip-solution II (strong strip)	0.5% SDS in dH_20

2.7. Enzymes

Klenow-Fragment (2U/µl) and 10 x reaction buffer	MBI Fermentas # EP0054
Proteinase K (Boehringer)	20 mg/ml solution in 10 mM Tris-HCl, pH 8.0 / 10 mM NaCl
RNase A (Boehringer)	10 mg/ml solution in 10 mM Tris-HCl, pH 8.0 / 15 mM NaCl 15 min boiled, for DNase inactivation, Store aliquots at –20°C
T*aq* DNA-Polymerase (5U/µl) and 10 x PCR-buffer	Gibco-BRL® # 18038
T4 DNA-Ligase (3U/µl) and respective 10x buffer	Promega # M1801

Restriction-Endonucleases

		New England Biolabs
A*fl* III	5U/µl	# R 0541S
Dde I	10U/µl	# R 0175S
F*sp* I	5U/µl	# R 0135S
H*ha* I	10U/µl	# R 0139S
Mlu I	10U/µl	# R0198
		MBI Fermentas
A*lu* I	10U/µl	# ER0011
Cla I	10U/µl	# ER0141
Dra I	10U/µl	# ER0221
*Eco*R V	10U/µl	# ER0301
Hae III	10U/µl	# ER0151
Hind III	10U/µl	# ER0501
Hpa II	10U/µl	# ER0511
Mbo I	10U/µl	# ER0811
Msp I	20U/µl	# ER0541
Taq I	10U/µl	# ER0671

All enzymes are supplied with the respective 10X reaction buffer.

2.8. cDNA synthesis

SuperScript™ First-Strand Synthesis System for RT-PCR Invitrogen #11904

2.9. Oligonucleotide/ Primer pairs (Interactiva – Thermo Hybaid)

Locus	Forward Primer (5'-3')	Reverse Primer (5'-3')	PCR (bp)	Tm (°C)
CDY1(p5)	CAGCCCGAGGGATAGAATCCATG	GCTGCTTTATATCCCAACAGAGG	450	55
CDY1 (p6)	TCCAATTGTACTGGAAGAATGTAAG	GAAGCGTTTGTGCCACTATAAAGTC	700	55
CDY1 (p7)	GGGCGAAAGCTGACAGCAA	GCTATGAAGATGGATTAGGGCA	396	70
BPY2	GTCCATAGTGGGCTTGCTGA	GCTTGTTCCACATATGGCACA	854	70
E1F1AY	CTCTGTAGTCAGCCTCTTC	GACTCCTTTCTGGCGGTTAC	90	60
XKRY	CACTCATGGAGAAGGGTAGG	GTCACACTCAGCCTCTTTAC	94	60
sY153	GCATCCTCATTTTATGTCCA	CAACCCAAAAGCACTGAGTA	140	57
SRY – sY14	GAATATTCCCGCTCTCCGGA	GCTGGTGCTCCATTCTTGAG	470	66.5
DAZ sY152	AAGACAGTCTGCCATGTTTCA	ACAGGAGGGTACTTAGCAGT	125	66.5
Y- DAZ3 3'	GCTAGAACCAAATATTCTGGAT	ACTCCTCAGACACTCTATTAGTT	260	66.5
5' ÚTR DAZ	ACGTAGCCTTCTTTCCAAAGCCGTAGG	CAGTGTTTCACCCACCACTTCTGG	3000	70
3'UTR DAZ	GAGCACCAGAGCCCTTTTGCTAT	CTTCAACAAAGACAAAAATTT	607	62
DAZ RRM3	CTGCTTAGCATTCTTCTATTATCTGT	ACGCAACTCTCACTGCTACATAGG	473	66
TPTE	ACAATAGAGACTCATCTGAGTAG	AGACTCGATAGCGGCATGGA	273	62
TPTE screen	AAAAGTATGTTAGAACGACTTTC	GAAAGTCGTTCTAACATACTTTT	1700	60
DAZ sY254	GGGTGTTACCAGAAGGCAAA	GAACCGTATCTACCAAAGCAGC	400	62/65
sY158	CTCAGAAGTCCTCCTAATAGTTCC	ACAGTGGTTTGTAGCGGGTA	231	60
DYZ2 sY159	TACATGTTATGTGCTATGCC	CACATTATATAATATGTATGTTGTC	461	58
DYZ1 sY160	TACGGGTCTCGAATGGAATA	TCATTGCATTCCTTTCCATT	236	62
sY167	GAAGGCATTAAGACATAAAGCA	ACTTTTGATATGAAAACAAACTTCA	117	62
DAZ intron3	CACTGCCCTAATCCTAGCACA	TCTTCTGGACATCCACGTCA	252	67
5'DAZ	CACAGGCACTCAGTAACTATCTC	CAGTGTTTCACCCACCACTTCTGG	710	66.5
THE-DAZ	GACAACACCACCGTACTCCA - sY579	AGGAAGGAAGTGGAGGGAAA	133	65
DAZ cDNA	CACGTTCAGTCTGCTGCAAA (ex6-2)	GAGGAGGAGGATTAACTACC (ex6)	412/907	60
DAZ cDNA	TAATTATCAGTCTGCTGCAAA (exI-2)	GAGGAGGAGGATTAACTACC (ex6)	412/907	60
DAZ cDNA	CACGTTCAGGCACACACTA (ex6-I)	GAGGAGGAGGATTAACTACC (ex6)	412/907	60
DAZ Line	GTGTATTCATGTATGCTACTG	CACTGGGAGATATACCTAAT	480	62
DAZ Prom.	ACGTAGCCTTCTTTCCAAAGCCGTAGG	CAGTGTTTCACCCACCACTTCTGG	3000	70
DAZex1-2	CGATAAGCATTTCTTCCCTCTACC	CGATAAGCATTTCTTCCCTCTACC	1300	66.8
DAZ cDNA	AGCAAAGGAGACATATCCTA (ex8-7G)	TGAGAGGATCCAACAGATTT	240	56
DAZ RRM	CTGCTTAGCATTCTTCTATTATCTGT	ACGCAACTCTCACTGCTACATAGG	1863	68
sY143	GCAGGATGAGAAGCAGGTAG	CCGTGTGCTGGAGACTAATC	311	62
DAZ1	GGAAGCTGCTTTGGTAGATAC	TAGGTTTCAGTGTTTGGATTCCG	1300	57/65
U3-sY139	TTCAGAGGAATCATGTGGGT	AATGTTTCATCACCATTATCCC	120/123	59
DAZ-SNV I	CAGTGTTTCACCCACCACTTCTGG	CACAGGCACTCAGTAACTATCTC	709	66.5
DAZ-SNV II	GGAAGCTGCTTTGGTAGATAC	GACATCCACGTCATTAACAAACG	674	65
DAZ-SNV III	GTGTGGCACATATGCCTATAAA	TTGGTACATCCAGATGCAGAT	301	63
DAZ-SNV IV	CTTCCTCATCTTTCTTGACTT	TTATTTATTCCTCAAAAAGGTG	660	57
DAZ-SNV V	TGGTTAATAAAGGGAAGGTGTTTT	TCTCCAGGACAGGAAAATCC	270	65
DAZ-SNV VI	GGGCCTAGTCTCTAGATCATT	GCTAGAACCAAATATTCTGGAT	431	64
AZFc-U3	ACTACCATTTCTGGAAGCCG – sY1192	CTCCCTTGGTTCATGCCATT	255	64
AZFc-U2-T2	TCATTTGTGTCCTTCTCTTGGA - sY1197	CTAAGCCAGGAACTTGCCAC	453	60
AZFc P1.1/1.2	CTGATTTGGAAACTCGTCCC – sY1190	CTGAAGAGATTGTGAAATGG	700	57
AZFc b4	ACCTAAGGGAACCCAGGAGA - sY1054	CGACACTTTTGGGAAGTTTCA	340	60
AZFc b4	CCGACTTCCACAATGGCT – sY1201	GGGAGAAAAGTTCTGCACG	677	64
Yellow SNV	ACAGTAGGGATGCCAATAAGATGG	TGGCTCAGGCTGAGCCTGAGGCAT	467	70
BPY2-SNV I	GTGTCTCTTGTGTCTGGACC	TCACTTCACCCGGGTACAGT	470	62
CSPG4L-SNV	GCCGAGTGGTGGCATGTGCA	TGCCAAGGAGCTGGAAGGTGA	380	66
GOLGA2LY	TGGGCCTGTTGCTTCTAGGGTT	ACAGGGAGGATGCTGCTGCTGTCACA	531	64
TTY-4 SNVI	GAGCCCTGCCTGCAGACA	GTATATGGCATAATTTCACCTG	540	58
DAZ1/4	CACGTTCAGTCTGCTGCAAA (ex6-2)	CCACAGATTTCTCCTTTGCT	1.2-1.7 kb	60
7A	CTCAGCACGTTCAGGCTTAT (ex6-7A)	CCACAGATTTCTCCTTTGCT	1400	65
7B	CTCAGCACGTTCAGGCTTAC (ex6-7B)	CCACAGATTTCTCCTTTGCT	650-1000	65

DAZ-short	TCAGTTCCATGTATACAATTAC	TCAAAACCAGCAACTTCCCATG	254	60
DAZBis	ATTTGATAGAAAGGAGTTTAAGT	TCACCATAATAATAACAAACAAC	674	58
DAZL	GATGTTAGGATGGATGAAACTG	AAGCCGGAGGTACAACATAG	490	60
DBX	GAAGCTACTAGAGGTTTCTAC	TCTCAACATCACTGAAACTTTC	420	64
CDNADAZ	Primer used for *DAZ* RT experiences	TCAAAACCAGCAACTTCCCATG	-	42

Table I - Primer pairs used in this work

2.10. DNA sources

2.10.1. Human PAC- and BAC-DNA clones

All BAC-*DAZ* clones belonging to the "RPCI-11 Human Male BAC" library are from RZPD - Deutsches Ressourcenzentrum für Genomforschung GmbH, Heubnerweg 6, D-14059 Berlin, www.rzpd.de - library nr. 753. The creator was Kazutoy Osoegawa, Minako Tateno (Shizuya *et al.* 1992).

The BAC-DAZ clones belonging to the "CTA/CTB Human Male BAC" library are from Research Genetics, 2130 Memorial Parkway SW, Huntsville, AL 35801, USA (Shizuya *et al.* 1992).

Name	Specificity	Acession nr.
RPCI-11 428D10	TPTE	AC019099
RPCI-11 557B9	TPTE	AC013734
CTB 352E14	DAZ 2	not sequenced
CTA 132B16	DAZ3	not sequenced
CTB 546E5	DAZ4	not sequenced
CTA 148I14	DAZ 1	AF164343
RPCI-11 290O3	DAZ 3/4[a]	AC010089
RPCI-11 140H23	DAZ 1/2[a]	AC053490
RPCI-11 539D10	DAZ 4[a]	AC006338
RPCI-11 226N11	DAZ 2	AC006990
RPCI-11 26D12	DAZ 4	AC006982
RPCI-11 263A15	DAZ 2/1[a]	AC007039
BAC K1020	DAZ 4 ?	not sequenced

Table II - BACs used in this work

The Human PAC-DAZ clones are from RZPD, library nr. 704, established by Pieter de Jong and P. Ioannou from the Roswell Park Cancer Institute, USA (Pierce *et al.* 1992).

Name	Clone	Specificity	Acession nr.
G19633	PAC 4C		not sequenced
G19660	PAC 12C		not sequenced
K07760	PAC 18F	DAZ 4	not sequenced
D061100	PAC 21C	DAZ 3	not sequenced
O131180	PAC 26D	DAZ 2	not sequenced

Table III - PACs used in this work

2.10.2. Human genomic DNA

All the FM (Fertile Men) samples were from normal fertile males.
All OZ (Oligozoospermic men) samples were from infertile males with normal karyotype and with sperm concentration < 10 millions/ml.

2.10.3. Vectors

2.10.3.1. pCR® 2.1-TOPO® vector

This vector was used for TOPO® cloning reactions using PCR products (*DAZ* 1.1, 1.8 and 1.9 probes), RT-PCR products and *DAZ* transcripts from cell lines. The map below (fig 9) shows the features of this vector and the sequence surrounding the TOPO® cloning site.

TOPO® TA Cloning kit, Invitrogen, # K 4500-01

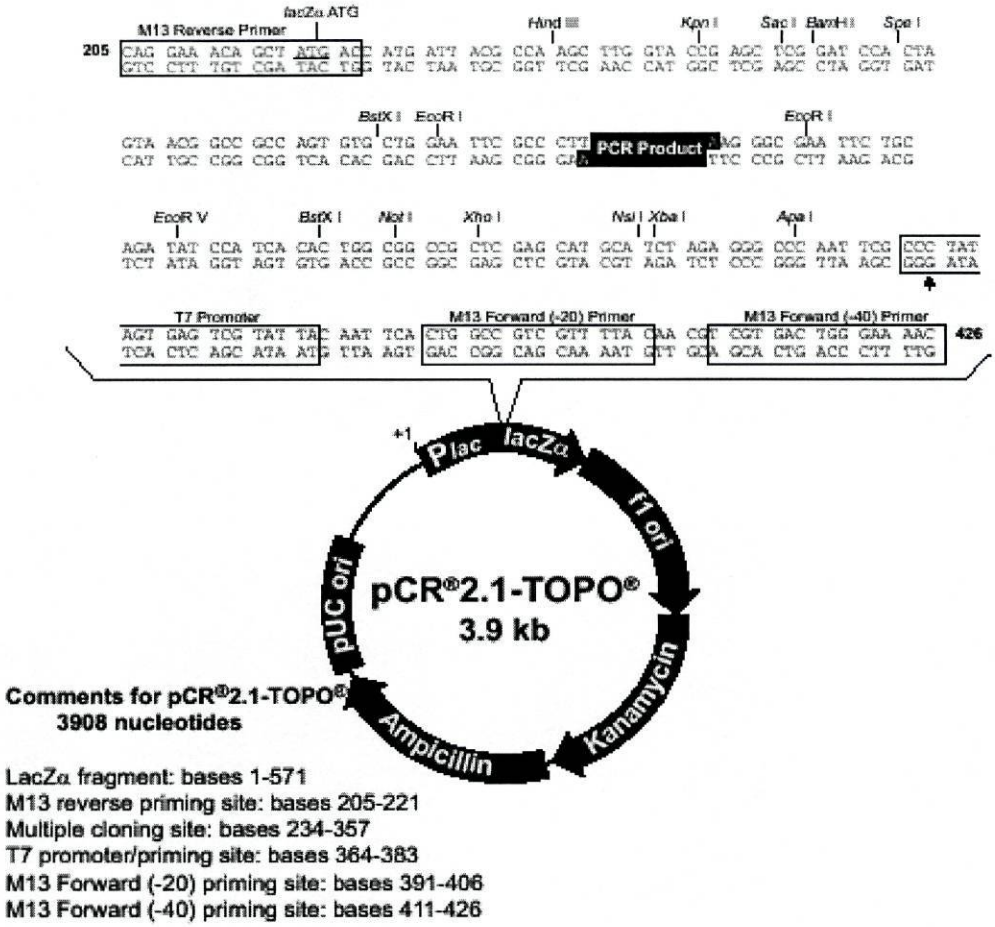
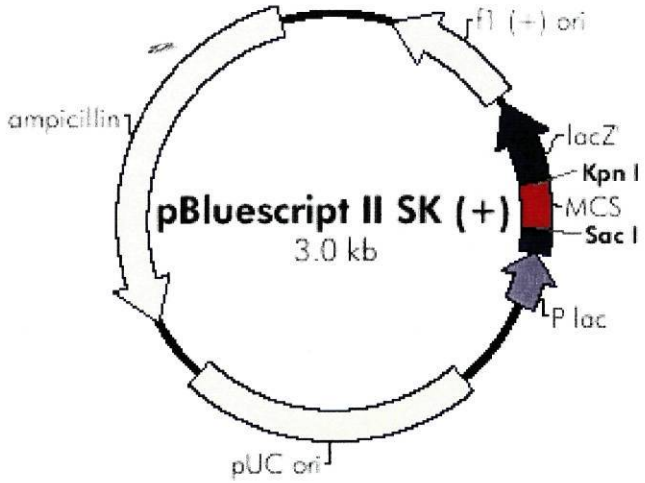


Figure 9 – Vector pCR® 2.1-TOPO®

2.10.3.2. pBluescript® II SK (+/-)

The inserts cloned in this vector (fig 10) were DAZL1-cDNAs, CT52Y, CT351Y and CT355Y, from the Human Testis Large Insert cDNA Library HL5503u (Clontech), containing the mRNA from normal, whole testes pooled from 25 Caucasians.



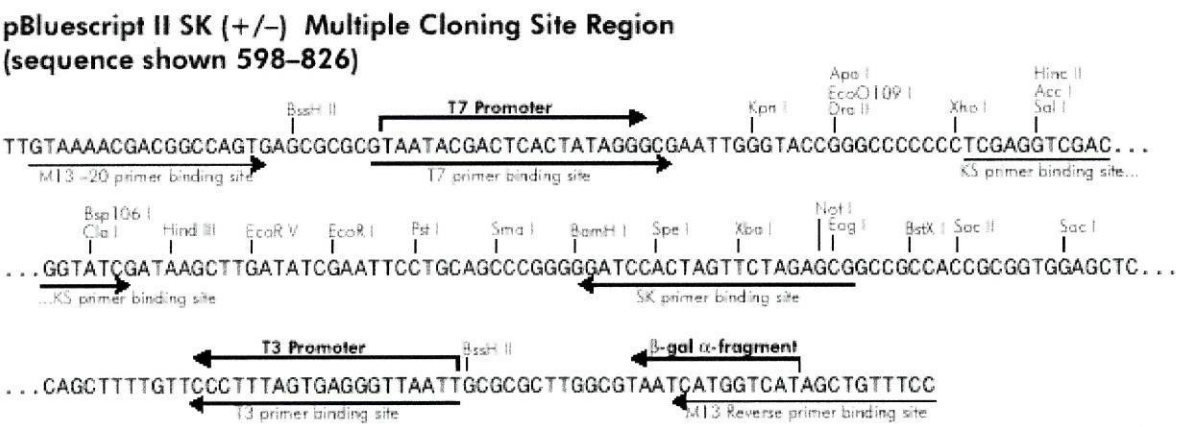


Figure 10 - Vector pBluescript® II SK (+/-)

2.10.3.3. pBACe3.6 vector

This pUC plasmid is currently used for BACs and PACs inserts in the range of 100-150 kb. The map below (fig. 11) shows the pBACe3.6 main features (Frengen *et al.* 1999). This vector has a cloning area to allow positive selection for inserts containing BAC clones through inclusion of the SACB II gene from P1 vector and disrupted the BamHI cloning site with a fragment containing a pUC plasmid (Pierce *et al.* 1992). All the BACs used in this work were cloned in this vector.

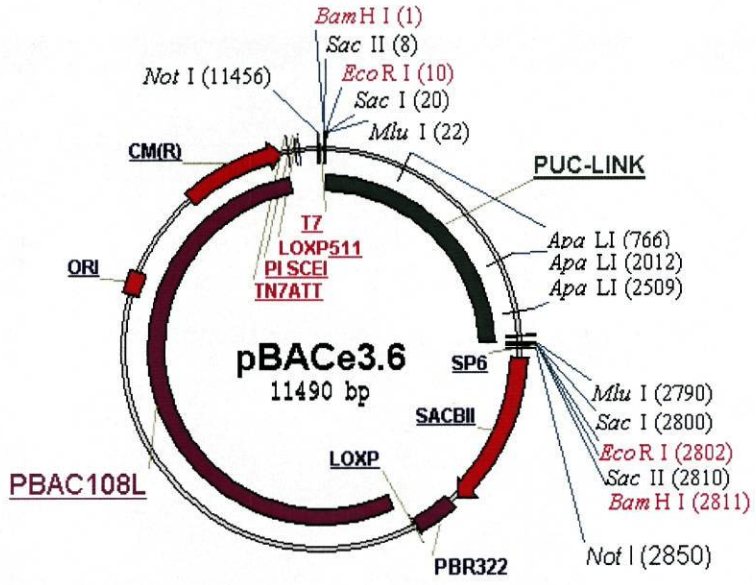


Figure 11 – Vector pBACe3.6

2.10.3.4. pCYPAC2 vector

The map below shows the pCYPAC2 main features (fig. 12). This vector has been constructed for the cloning of large DNA fragments following electroporation, by inserting a pUC plasmid into the BamHI cloning site (Pierce *et al.* 1992). All the PACs used in this work were cloned in this vector.

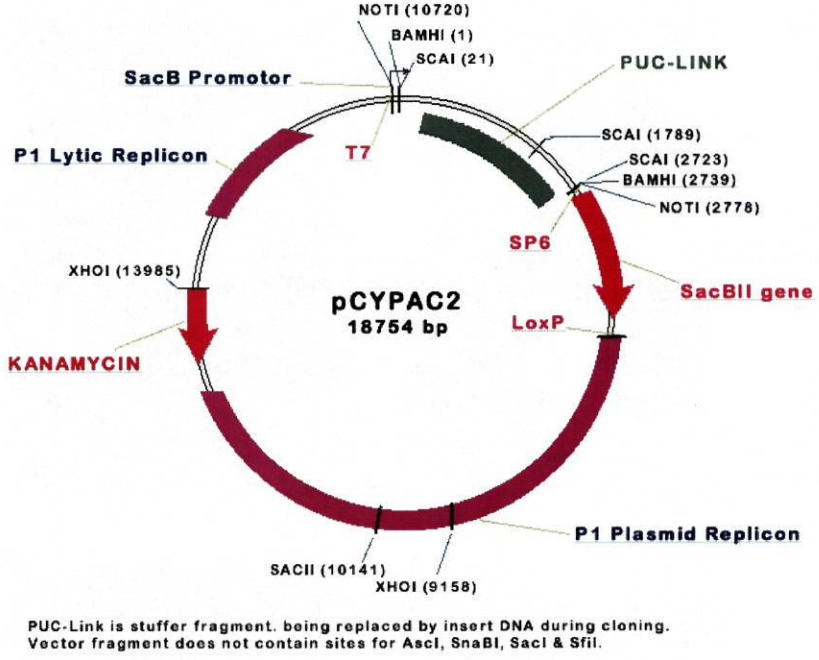


Figure 12 - Vector pCYPAC2

2.10.4. Probes

Name	Type	Specificity	Size
CT52Y(EcoRI-AvaII)	cDNA	DAZ 7 variants	800 bp
CT351 A23 (Hind III)	cDNA	DAZ (ex1-ex7)	2 kb
DAZ1.1	genomic	DAZ (ex1-ex2)	1.4 kb
DAZ1.8	genomic	RRM (incl 7G)	1.8 kb
49f	genomic	DAZ (ex 4,5,6,7B)	2.8 kb
DAZ1.9	genomic	Putative promoter	1.9 kb

Table IV - Probes used in this work

2.11. Cell culture

RPMI 1640 (without L-Glutamine)	Gibco-BRL® # 31870
Dulbecco´s Modified Eagle Medium (DMEM) (1000 mg/l glucose)	Gibco-BRL® # 31885
Fetal Bovine Serum (FBS) (heat inactivation at 56°C for 30 min. and sterile filter)	Gibco-BRL® # 12476
L-Glutamine 200 mM	Gibco-BRL® # 15039
Antibiotic-Antimycotic solution (Pen-Strep-Amp)	Gibco-BRL® # 15240-062
Trypsin-EDTA solution 10 X	Gibco-BRL® # 35400
Non-Essential-Aminoacids (NEA)	Gibco-BRL® # 11140
2ß-Mercaptoethanol (2ß-ME)	Gibco-BRL® # 31350
5'-Azacytidine (4-amino-1-ß-D-ribofuranosyl-5-triazin-2[1H]-one	Sigma # A2385
Trichostatin A (4,6-Dimethyl-7-(p.dimethylaminophenil) -7-oxohepta-2,4,dienohydroxamic acid)	Sigma # T8552
Na-bisulfite	Sigma # S-9000
Hydroquinone	Sigma # H-9003
Wizard® DNA Clean-Up system	Promega # A7280

2.12. Photographic material

Autoradiography

X-ray films	X-OMAT AR (XAR-5), Kodak
Developer	G153 Developer, Agfa
Fixative	G354 RapidFixer, Agfa

2.13. *In silico* - Sequence analysis

2.13.1. Databank

Data Bank from National Center for Biotechnology Information (NCBI),
National Institutes of Health, National Library of Medicine, National Institutes of Health (NIH), Washington, USA
http://www.ncbi.nlm.nih.gov

2.13.2. Software for sequence analysis

HUSAR (Heidelberg Unix Sequenz Analysis Resources)
The Biocomputing Service Group, DKFZ Heidelberg, Germany
http://genome-dkfz-heidelberg.de
http://genius.embnet.dkfz-heidelberg.de/menu/w2h/w2hdfz/index.html
Clustal and GAP - Comparing homologies among sequences
Map and Mapsort - Restrictions analysis and translation

BLAST (Basic Local Alignment Search Tool, Altschul et al. 1990, 1997)
National Center for Biotechnology Information, Bethesda, Maryland, USA
http://www.ncbi.nlm.nih.gov/BLAST/

NIX (Identity unknown nucleic sequences)
UK Human Genome Mapping Project Resource Centre, MRC, UK
http://www.hgmp.mrc.ac.uk/menu-bin/Nix/

Repbase (Genetic Information Research Institut)
http://www.girinst.org/

DNA-fragment analysis

Dnafrag Version 3.03 (Schaffer *et al.* 1981)

Programmed by John Nash

Copyright © National Research Council of Canada, 1991.

2.14. Strains and cell lines

2.14.1. Bacterial strains

The bacterial strains were obtained by transforming the E. coli hosts given below with the corresponding recombinant plasmids:

- E. coli TOP10F' One Shot® Chemically Competent cells - F´ *lacI*^{q} Tn*10* *(Tet*^{R}*) mcr*A Δ(*mrr-hsdRMS-mcrBC)* ϕ*80lacZ*ΔM15 Δ*lac*X74 *rec*A1 *deo*R *ara* D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nupG.* This strain over expresses the Lac repressor (*lacI*^{q} gene). This strain contains the F episome and can be used for single-strand rescue of plasmid DNA containing an f1 origin - Invitrogen # K4550

- DH5∝™-T1^R - F- ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(r_k^-,m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *ton*A. This strain is used for general cloning and blue/white screening without IPTG. Strain is resistance to bacteriophage T1 - Invitrogen # K4520.

2.14.2. Human cell lines

The **NT2/D1** cell line is a human germ-cell cancer cell line initially obtained from Prof. P. Andrews, Sheffield, UK, growing as a monolayer in DMEM supplemented with 10% FBS, 0.1 mM 2ß-ME and 1% Pen-Strep-Amp (Andrews 1984).

The **Ep2102** cell line is a human germ-cell cancer cell line initially obtained from Prof. P. Andrews, Sheffield, UK, growing in DMEM supplemented with 10% FBS and 1% Pen-Strep-Amp (Andrews 1984).

The **Du-145** cell line is a human prostate cancer cell line obtained from DKFZ (Deutsches Krebsforschungszentrum), Heidelberg, Germany, growing as a monolayer in DMEM supplemented with 10% FBS and 1% Pen-Strep-Amp.

The **LNCaP** cell line is a human prostate cancer cell line obtained from DKFZ (Deutsches Krebsforschungszentrum), Heidelberg, Germany, growing as a monolayer in RPMI 1640 supplemented with 10% FBS, 1% L-Glutamine, 1% Pen-Strep-Amp and 1% NEA (Webber *et al.* 1996).

2.15. Sequencing

BigDye™ Terminator Cycle Sequencing v2.0 kit	Applied Biosystems # 4314414
MicroSpin™ S-400 HR columns	Amersham Pharmacia Biotech
ABI PRISM® 310 Genetic Analyser	Applied Biosystems

Methods

3.1. Microbiology techniques

3.1.1. Bacterial cultures

The bacteria can be cultivated in agar or in liquid medium. The plate cultures containing isolated colonies can be stored at 4°C for long time.

Culture in agar plates

- Add 15 g BACTO-Agar per litre LB medium and autoclave
- After cooling down the medium (ca. 50°C) add the appropriate antibiotic and poor the medium in Petri dishes
- Spread the bacteria on the plates and incubate for 12 to 20 hours at 37°C

Culture in liquid medium

- The liquid cultures can be performed in volumes from 5 to 500 ml
- After autoclaving, cool the medium down (ca. 50°C) and add the appropriate antibiotic
- Inoculate the medium with a single colony (from an agar plate culture) or 1:100 v/v dilution from a main culture
- Incubate for 12 to 20 hours at 37°C with shaking
- The bacterial concentration can be measured by optical density at 600 nm and the concentration is calculated by the formula: $1OD_{600} = 5,8 \times 10^8$ bacteria/ml

3.1.2. Bacterial stocks (Glycerol stocks)

The long time storage (years) of bacteria is done at -70°C in 15% glycerol.

- Mix 0.5 ml overnight bacterial culture with 0.5 ml Glycerol 87% solution
- For long time storage, store the tubes at -70°C

3.1.3. Screening of genomic libraries

The PAC- and BAC-DNA libraries from RZPD are the so-called XYZ genomic libraries and can be screened by a two-step PCR procedure. The BAC-DNA library contains 423 936 clones, grouped in 23 primary pools consisting of purified DNA from all clones from 8 microtiterplates of the genomic library. The identification of the desired clones is done by testing first the 23 primary pools. Once a positive signal with a primary pool is obtained, the corresponding secondary pool is sent. The secondary pools consist of individually pooled microtiterplates from the positive primary pool, as well as pooled rows and columns from all 8 microtiterplates of the positive primary pool. Each secondary

pool consists of 48 pools maintained in 384-well microtiterplates. The PCR screening of each secondary pool should result in 3 positive signals: the positive plate, the positive row and the positive column. With the combination of these 3 coordinates, the exact clone can be identified and acquired from the Resource Center (RZPD).

For the PCR screening, the lyophilised BAC-DNA should be resuspended in 5μl dH_2O from the microtiterplate wells and incubated at room temperature with shaking for 15 min. 1 μl from each suspension should be used for PCR reaction. For this work, the pools from RPCI-11 Human Male BAC library were screened with the *DAZ* marker sY254 (GenBank, ac. nr. G38349) to find the *DAZ*-BAC clones.

3.2. Agarose gel electrophoresis

The DNA and RNA can be separated by molecular weight by electrophoresis in an agarose gel electrophoresis. Agarose is a polymer of D-Galactose and 3,6-anhydro-L-Galactose isolated from a red seaweed. Agarose gel is prepared by dissolving the powder in 1x TAE buffer. The DNA samples are mixed with the appropriate volume of loading buffer and the gel is loaded using 10 to 25ul final solution. Because DNA in solution is negatively charged, the DNA migrates under the influence of an electric field from the cathode to the anode. The DNA migration depends on the DNA size and conformation, buffer and agarose concentrations.

The migration of linear, double strain DNA molecules in agarose matrix is invertely proportional to the size logarithms. The super-coiled and circular plasmidic DNA migrates faster than the corresponding relaxed molecule.

The relative mobility is first depending on the agarose concentration on the gel. On table V are listed the optimal agarose gel concentrations for separating different sizes of DNA.

Agarose concentration [%w/v]	DNA sizes [kb]
0.3	5-60
0.6	1-20
0.7	0.8 – 10
0.9	0.5 – 7
1.2	0.4 – 6
1.5	0.2 – 3
2.0	0.1 – 2

Table V - Agarose concentration vs DNA fragment sizes.

For RNA, samples (5μg) were denatured in loading buffer containing 50% formamide, 18% deionised formaldehyde and 0.02M MOPS, heated at 68°C for 10 min., and then separated through a 1.2% agarose denaturing gel containing 18% formaldehyde and 0.02M MOPS. The running gel buffer was 0.02M MOPS, pH 7.0. The electrophoresis was performed for 3h at 60V.

For DNA and RNA visualisation, stain the agarose gel with 0,001% Ethidium bromid solution. Ethidium bromide intercalates between the bases and, under U.V. light (366 nm), emits light at visible wavelength (590 nm), allowing the nucleic acids to be seen. For documentation, a picture is taken from the gel on a U.V.-transilluminator.

More precise size of each fragment on the gel can be estimated running, on the same conditions, a molecular weight marker (DNA or RNA-Ladder).

3.3. DNA isolation

3.3.1. From blood

For DNA extraction from peripheral blood (Miller *et al.* 1988), the main steps are the lysis of erythrocytes, disruption of cellular membranes, protein denaturation and DNA precipitation.

❖ ***Lysis***

- Mix 10 ml EDTA-blood with 30 ml lysis buffer (ice cold) and incubate on ice for 30 min.
- Centrifuge at 1500 rpm for 10 min. and wash the pellet with 10 ml lysis buffer to obtain a white pellet
- Resuspend the pellet in 10 ml SE buffer (Nucleus-Lysis buffer), centrifuge and resuspend in 5 ml SE buffer (ice cold)
- Add 40 µl Proteinase K (10 mg/ml in SE buffer) and 250 µl of 20% SDS
- Mix gently and incubate O.N. at 37°C (the solution should be clear)

❖ ***Salt precipitation***

- Fill up with SE buffer to a final volume of 15 ml
- Precipitate cells debris with 3,3 ml saturated NaCl solution, mix by vortex and centrifuge at 3000 rpm for 15 min
- Discard the pellet and precipitate genomic DNA from the supernatant with 0.1 volume 3 M Na-Ac pH 7.0 and 1 volume Isopropanol
- Invert the tube gently until the DNA strand appears
- Hook the DNA and transfer to an eppendorf tube
- Wash DNA with 500 µl 80% Ethanol and air-dry the pellet
- Resuspend the DNA in TE buffer and put it in a rotor at 4°C

3.3.2. Plasmid preparation from bacterial cultures

3.3.2.1. DNA Mini-preps

The Plasmid-DNA (till 20 µg) preparation was done using the *Plasmid Mini Kit* (Qiagen®, Hilden, Deutschland). Qiagen® plasmide purification protocols are based on a modified alkaline lysis procedure (Birnboim & Doly, 1979), followed by binding of plasmid DNA to Qiagen® anion-exchange resin, under appropriate low-salt and pH conditions. RNA, proteins and low molecular weight impurities are removed by a medium-salt wash. Plasmid DNA is eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation. These protocols are optimised for use with cultures grown in standard Luria Bertani (LB) medium to a cell density of ~$1x10^9$ cells/ml.

❖ ***Start culture***

- Add 10 µl of Ampicillin stock solution to 5 ml LB medium
- Pick a single colony from a streaked selective plate and inoculate a culture of 5 ml LB medium
- Incubate over night at 37°C with vigorous shaking (~ 300 rpm)
- Harvest the bacterial cells by centrifugation at 6000 x g for 15 min at 4°C

❖ ***Bacterial lysis***

- Resuspend the bacterial pellet in 0.3 ml buffer P1 (+RNase A)
- Add 0.3 ml buffer P2, mix gently but thoroughly by inverting 4-6 times and incubate at room temperature for 5 min.

- Add 0.3 ml of chilled (4°C) buffer P3, mix immediately but gently by inverting and incubate on ice for 5 min.
- Centrifuge at 11000x g for 10 min. and change supernatant to a new Eppendorf tube

❖ ***Cleaning the Plasmid-DNA in a anion-exchange column***

- Equilibrate a Qiagen®-tip 20 by applying 1 ml buffer QBT and allow the column to empty by gravity flow
- Apply the supernatant (2 steps below) to the Qiagen®-tip and allow it to enter the resin by gravity flow
- Wash the Qiagen®-tip with 4 x 1 ml buffer QC, to remove all contaminations from the plasmid preparation
- Elute Plasmid-DNA with 0.8 ml buffer QF

❖ ***Plasmid-DNA preparation***

- Precipitate the Plasmid-DNA by adding 0.56 ml (0,7 Volume) Isopropanol, mix and centrifuge at 11000 rpm for 30 min. Carefully decant the supernatant
- Wash DNA pellet with 1 ml 70% Ethanol and centrifuge at 11000 rpm for 15 min., decant the supernatant without disturbing the pellet
- Air-dry DNA pellet for 5-10 min and resuspend the DNA in 30 µl TE buffer

3.3.3. BAC-DNA preparation

The genomic DNA extraction from large constructs (up to 250 kb) such as BACs (Bacterial Artificial Chromosomes) and PACs (Phage Artificial Chromosomes), was done using the QIAGEN® *Large-Construct Kit* (QIAGEN®, Hilden, Deutschland). The QIAGEN® Large-Construct kit protocols are based on a modified alkaline lysis procedure and an integrated ATP-dependent exonuclease digestion. This step ensures selective removal of contaminating genomic DNA as well as nicked or damaged large-construct DNA. Following the digestion step, large-construct DNA is bound QIAGEN® resin. RNA, proteins and low molecular weight impurities are removed by a medium-salt wash. Large-construct DNA is eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation.

❖ ***Start culture***

- Add 5 µl of Chloramphenicol stock solution (12,5 mg/ml) to 5 ml LB medium
- Pick a single colony from a streaked selective plate and inoculate a starter culture of 5 ml LB medium/Chloramphenicol
- Incubate 8 h at 37°C with vigorous shaking (~ 300 rpm)

❖ ***Overnight culture***

- Add 500 µl of Chloramphenicol stock solution in 500 ml LB medium

- Dilute 1 ml of the starter culture in 500 ml selective LB medium
- Incubate 12 to 16 h at 37°C with vigorous shaking (~ 300 rpm)
- Harvest the bacterial cells by centrifugation at 6000 x g for 15 min. at 4°C

❖ ***Bacterial lysis***

- Resuspend the bacterial pellet in 20 ml of buffer P1 (+Rnase A, 100 µg/ml)
- Add 20 ml of buffer P2, mix gently but thoroughly and incubate at room temperature for 5 min.
- Add 20 ml of chilled (4°C) buffer P3, mix immediately and incubate on ice for 10 min.
- Centrifuge in a Sorvall Rotor at 9000 rpm for 30 min. at 4°C
- Remove supernatant (containing BAC/PAC DNA) promptly to a 50 ml Falcon tube
- Filter the lysate through a filter pre-wetted with dH_2O

❖ ***DNA precipitation***

- Precipitate DNA by adding 0,6 volumes (ca. 36 ml) isopropanol to the lysate, shortly mix and centrifuge at 3500 rpm for 60 min. at 4°C
- Wash DNA pellet with 5 ml 70% Ethanol and centrifuge at 3500 rpm for 30 min.
- After discard the supernatant, air-dry the pellet and re-dissolve the DNA in 9.5 ml of buffer EX.

❖ ***Exonuclease digestion***

- Add 200 µl of ATP-dependent exonuclease and 300 µl ATP stock solution to the dissolved DNA, mix gently and incubate at 37°C for 60 min.

❖ ***Cleaning the BAC-DNA through a column***

- Equilibrate a QIAGEN®-tip 500 with 10 ml buffer QBT
- After the genomic DNA digestion with exonuclease, add 10 ml of buffer QS, apply the whole sample to the QIAGEN®-tip and allow it to enter the resin by gravity flow
- Wash the QIAGEN®-tip with 2 x 30 ml buffer QC
- Elute BAC-DNA with 15 ml buffer QF, pre-warmed to 65°C

❖ ***BAC-DNA precipitation***

- Add 10,5 ml (0,7 volumes) isopropanol to the eluted DNA
- Mix and centrifuge at 3500 rpm for 60 min at 4°C
- After decanting the supernatant, wash DNA pellet with 5 ml 70% Ethanol and centrifuge at 3500 rpm for 30 min.
- Air-dry the pellet and re-dissolve the DNA in 30 µl TE buffer

3.4. RNA isolation

For this purpose the QIAamp® RNA blood mini kit from QIAGEN® was used. This kit provides a fast and easy method for the preparation of total cellular RNA from human whole blood. In addition, this kit can be used to purify total RNA from tissues and cultured cells.

3.4.1. From blood

QIAamp® spin columns technology combines the selective binding properties of a silica-gel based membrane with the speed and convenience of microspin technology. A high-salt buffering system allows RNAs longer than 200 bases to bind to the QIAamp® membrane. During this procedure, the main steps are the lysis of erythrocytes followed by the lysis of leukocytes using highly denaturing conditions to inactivate RNases, allowing the isolation of intact RNA. RNA is then bound to the silica-gel membrane and the contaminations are washed away. The RNA is finally eluted from the spin column using RNase-free water.

❖ ***Lysis***

- Mix 1 volume of EDTA-blood with 5 volumes of buffer EL and incubate on ice for 15 min.
- Centrifuge at 1500 rpm for 10 min. at 4°C and discard the supernatant. Wash 2 or 3 times with buffer EL to obtain a white pellet
- Add 350 or 600µl buffer RLT (depending on leukocytes number) and vortex or pipette to mix

❖ ***Homogenization***

- Pipet lysate into a QIAshredder® spin column and centrifuge for 2 min. at maximum speed to homogenize.
- Discard QIAshredder® spin column and save homogenized lysate
- Add 1 volume of 70% ethanol to the lysate and mix by pipetting

❖ ***RNA cleaning (including DNase treatment)***

- Pipet sample, including any precipitate which may have formed, into a Rneasy® Mini column
- Centrifuge for 1 min. at 10000 rpm and transfer the column to a new 2 ml tube
- Wash with 350µl buffer RW1 and centrifuge for 1 min. at 10000 rpm
- Add 10µl DNase I stock solution to 70µl buffer RDD and mix by inverting the tube
- Apply the DNase incubation mix to the column and incubate at room temperature for 15 min.
- To wash, pipet 350µl buffer RW1 and centrifuge for 1 min. at 10000 rpm
- Pipet 500µl of buffer RPE into the column and centrifuge for 1 min. at 10000 rpm
- To remove residual ethanol, add 500µl of buffer RPE and centrifuge for 2 min. at maximum speed
- Transfer spin-column to a new 1.5ml tube and elute RNA with 30µl RNase-free water. Centrifuge for 1 min. at 10000 rpm
- Repeat the elution step to a new 1.5ml tube if the expected RNA yield is >30 µg

3.4.2. From cultured cells

Cell pellets in buffer RLT can be stored at -70°C for several months. To process frozen lysates, thaw and incubate at 37°C for 10 min. to ensure that all the salt has dissolved.

❖ ***Harvest cells***

- Cells growing in a monolayer can be trypsinized and collected as a cell pellet prior to lysis
- Discard medium, wash cells 2x with PBS, aspirate and add the trypsin solution
- After cells detach from the flask, add PBS, transfer to a tube and pellet by centrifugation at 1200 rpm for 5 min.
- Disrupt cells by adding 350 or 600µl buffer RLT (depending on pellet cells number), vortex or pipette to mix. No cell clumps should be visible before proceeding to the homogenisation steps

❖ ***Homogenization and following steps***

- The protocol used for the following steps is the same as for RNA isolation from blood (see above)

3.5. Genomic DNA Blot experiments

3.5.1. Sample preparation

Briefly, the DNA samples were digested with different restriction enzymes, restricted DNA fragments separated on agarose gel by electrophoresis and transferred to nylon membranes by blotting according to Southern (Southern, 1975).

❖ ***Samples preparation***

- Prepare 8 µg genomic DNA from each sample. If the sample is too old or to viscous incubate at 65°C for 5 min. before pipetting it
- Precipitate with 1/10 Na Ac and 2V absolute ethanol
- Incubate at -20°C for 30 min. and centrifuge 30 min. at 4°C
- Dry pellet and resuspend on appropriate volume with TE buffer (10ul for a medium gel; 12.4ul for a big gel)

❖ ***Endonuclease restrictions and gel running***

- Add to the **genomic DNA**, 1/10 volume spermidine (25 mM), 1x RE buffer and 50 U RE in a final volume of 20 µl (for **plasmidic DNA**, no need of spermidine)
- Incubate O.N. at 37°C (for Taq I incubation at 65°C for 3 h)
- Add 2 µl of 6x loading buffer, mix and incubate at 65°C for 10 min.
- Load all volume (20ul) on the gel (0,8% gel with LE-Seakem® agarose)
- Run gel at 200V for 10-15 min and then 30-75V O.N. or until the orange dye is at the end of the gel
- Stain the gel in Ethidium bromide and take a real size picture

3.5.2. Capillary Blot

For the DNA transfer to a nylon membrane, after the electrophoresis the DNA should be depurinated with HCl, denatured with NaOH and neutralised with Tris buffer. After the DNA transfer to the nylon membrane, the membrane is UV-crosslinked (254 nm) and baked to bind DNA permanently to the membrane.

❖ ***Gel treatment:***

- Depurination - 0.25M HCl (freshly prepared) for 15 min. (the orange colour should change to grey)
- Denaturation - 0.5M NaOH / 1.5M NaCl for 2X 45 min. (the blue colour will appear again)
- Neutralisation - 1M Tris (pH 7.5) / 1.5M NaCl

❖ ***DNA transfer***

- Cut the nylon membrane (Hybond N+) and wet in hot water, followed by equilibration step on 2XSSC for 10 min.
- Mount the blot as shown in figure 1MM
- Transfer O.N. in 20X SSC (3M NaCl/ 0.3M Natrium citrat/pH7.0 with HCl)
- Dismount the blot, UV cross-link the membrane and bake for 2h at 80°C to bind permanently the DNA to the membrane
- Store at R.T. or use to hybridise with a specific probe
- Stain the gel in EtBr to check if the transfer was complete

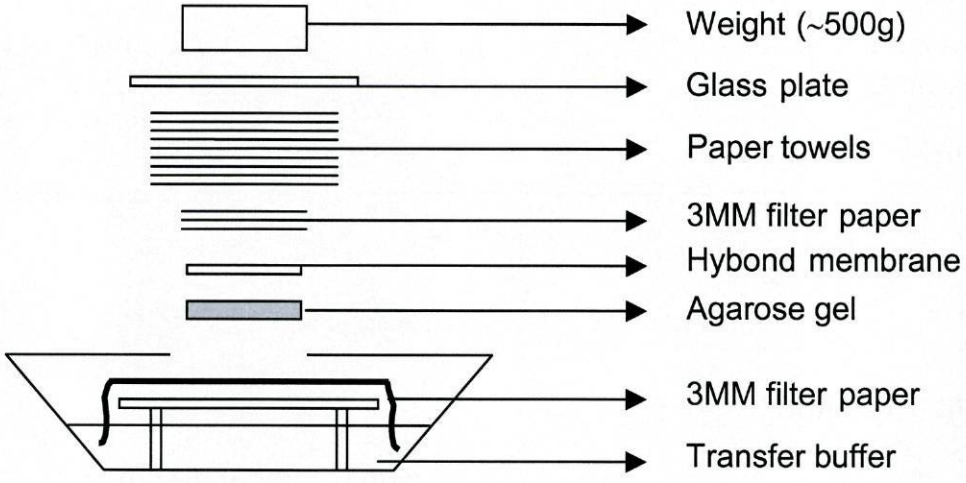


Figure 13 – Capillary blots by Southern (according to Southern, 1975).

3.5.3. Vacuum-Blot

The DNA transfer from an agarose gel to a membrane can also be done by vacuum with a vacuum blot system from Amersham Pharmacia Biotech.

❖ ***Building the vacuum blot***

- Cut the nylon membrane and treat as for capillary blot (wet in hot water and equilibrate on 2x SSC) and put in the vacuum apparatus
- Cut and adjust a plastic mask to the size of the gel
- Add the agarose gel, after stained with EtBr (seal with Bromophenol blue gel)

❖ ***DNA transfer***

- Adjust the vacuum to 55 cm^3 and treat the gel as follows:
  Depurination - 0.25 M HCl for 2 x 15 min.
  Denaturation - 0.5 M NaOH/1.5 M NaCl for 1 x 5 min., followed by 2 x 30 min

❖ ***Dismount the blot***

- Wash the membrane shortly in 2 x SSC
- UV- crosslink (1.2 J/cm^2 for 30 sec.) the membrane and bake for 2h at 80°C to bind permanently the DNA to the membrane
- Store at R.T. or use to hybridise with a specific probe
- Stain the gel in EtBr to check if the transfer was complete

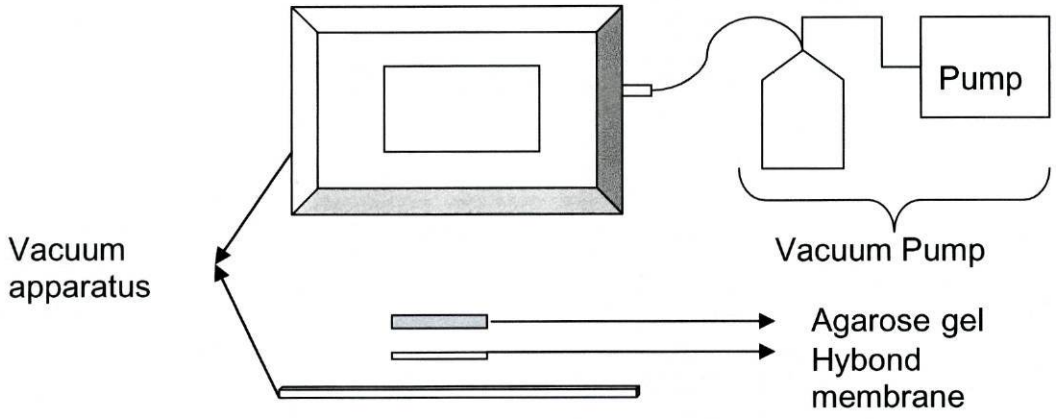


Figure 14 – Vacuum blots scheme.

3.6. Radioactive hybridisation

The blots were pre-hybridised for 3 hours in Church buffer at 65°C, followed by hybridisation at 65°C overnight with a specific probe. The DNA sequence to be used as a probe is labelled with radioactivity ($\propto$ ^{32}P-dCTP). Blots were washed under stringent conditions (2 x SSC / 0.1% SDS at 65°C). Membranes were exposed to auto-radiography at -70°C with an intensifying screen for 1-3 days.

3.6.1. Radioactive labelling by Random Priming

The complementary strand of DNA is synthesised by a DNA polymerase I (Klenow-Fragment) from a template DNA using hexanucleotides mix as a random primer. The nucleotide labelled ([^{32}P] dCTP) is incorporated during the new strain synthesis.

- Add 20-100 ng DNA (volume variable) to 4x a Hexamer-Oligonucleotide (80-400 ng)
- Complete the volume to 11µl with dH_2O
- Denature at 95°C for 5 min., centrifuge shortly and cool down on ice
- On ice, add 2µl 10 x Reaction buffer, 1µl 33 mM dATG mix and 5µl α ^{32}P-dCTP (50 µCi)
- The DNA synthesis starts with the addition of 1µl Klenow-Fragment (2 U/µl), followed by incubation at 37°C for 3 h in a hot block

3.6.2. Probe cleaning

In order to have a specific hybridisation and to reduce the background it is important to remove the unincorporated radioactive nucleotides. This is normally done by using a Sephadex®-G50-column (NICK™ Column, Amersham Pharmacia Biotech). The use of these columns is not recommended for purification of oligonucleotides shorter than 20bp.

- Equilibrate the gel from the column with 3 x 3ml dH_2O
- Allow the dH_2O to completely enter the gel bed
- Add 80µl dH_2O to the labelled probe (20 µl) and load the column with all volume (100µl)
- Add 300 µl dH_2O and allow to enter into the gel bed
- Discard the first sample collection (400 µl fraction)
- Elute the purified sample (elution containing the probe) after adding 2x 400 µl dH_2O

3.6.3. DNA-blot hybridisation

After labelling and cleaning the probe, but before the hybridisation, the DNA-probe should be denatured.

- To avoid non-specific hybridisation, add 200µl Salmon sperm DNA (1 mg/ml) to 800µl probe preparation in order to block the repetitive DNA
- Add 31µl of 10M NaOH solution (final concentration 0,3M) to the cleaned DNA-probe (1ml) and incubate at room temperature for 10 min.
- Stop probe denaturation by neutralisation with 31µl of 10M acetic acid solution (final concentration 0,3M) and proceed to the hybridisation

The pre- and hybridisation are done at 65°C, in rotating tubes in a hybridisation oven.

❖ ***Pre-hybridisation***

- Add 10ml Church buffer to the nylon membrane and incubate at 65°C for 3 h

❖ ***Hybridisation***

- Add the denatured DNA-probe to 5 ml Church buffer pre-warmed at 65°C
- Hybridise the membrane for 8-12h (overnight) at 65°C

❖ ***Post-hybridisation washes (to remove non-specific binding)***

- The membrane washes were done with 2x SSC / 0.1% SDS at 65°C, as follows:

  1 × 20 min. in the glass tubes in the hybridisation oven
  2 × 15 min. at 65 °C in the water bath (with shaking)

  After each washing step, the background is measured with a Geiger counter. The washing steps should stop when the background is near zero. If necessary, more stringent washing steps should be performed with 1x SSC / 0.1% SDS at 65 °C, 0.1x SSC / 0.1% SDS, etc.

❖ ***Membrane exposition after hybridisation***

- The membrane should be covered by plastic folio and exposed to auto-radiography at -80°C in a cassette with intensifying screens
- The exposure time is variable and depending on the type of blot: genomic blot - at least overnight exposition; plasmidic blot - 1h - 5 h

The probes used during this work were produced from cDNA transcripts (cDNA probes) cloned in different vectors or genomic probes produced by PCR amplification and cloned on TOPO® vector.

3.7. Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) is an enzymatic method for specific exponential amplification of a DNA fragment *in vitro* (Saiki *et al.* 1985).

3.7.1. Standard PCR

The standard PCR protocol was performed in 25 µl total volume as follows:

2.5 µl	10 x PCR buffer
1.25 µl	50 mM $MgCl_2^*$
0.3 µl	25 mM dNTPs
1.0 µl	10 µM Primer mix (for/rev)
0.3 µl	5 U/µl Taq Polymerase
1.0 µl	100 ng DNA template
25.0 µl	final volume with dH_2O

* Changes on Magnesium concentration can improve the amplification efficiency (Henegariu *et al.* 1997)

3.7.2. Standard PCR programme

The PCRs were performed in a Biometra T1 Personal Cycler or in a Biometra T-Gradient Cycler. A standard PCR programme (see below) was used, with the annealing temperature depending on the primers used and listed on table I:

PCR standard programme:

First denaturation	94°C	4 min	
Denaturation	94°C	1 min	⎤
Annealing	variable*	1 min	35 cycles
Elongation	72°C	1 min	⎦
Final extension	72°C	4 min	

* Primer specific temperature calculated by the formula (Maniatis *et al.* 2001):

Annealing temperature [°C] = 4 x (G +C) + 2 x (A+T)

3.7.3. *DAZ* SNV/STS analyses

On table 2MM are described the special PCR conditions used during this work for *DAZ*-SNVs and STS analyses.

SNV / STS	$MgCl_2$	Annealing Temperature
SNV I	2.0 mM	66,5 °C
SNV II	2.0 mM	65 °C
SNV III	2.5 mM	63 °C
SNV IV	2.5 mM	57 °C
SNV V	2.0 mM	65 °C
SNV VI	2.0 mM	64 °C
Y-RRM3	2.0 mM	66 °C
SY254	2.0 mM	65 °C

TRIPLEX

STS	$MgCl_2$	Annealing Temperature	Primer mix concentration
sY14 (*SRY*)	2,5 mM	66 °C	5 µM
Y-*DAZ*3			20 µM
sY152			10 µM

Table VI – PCR conditions for *DAZ*-SNVs/STS.

On the Triplex PCR, the marker sY14 is an internal positive control for the *SRY*-locus and the sY152 is a marker for *DAZ1/DAZ4* gene copies (Saxena *et al.* 2000). PCR products (10 µl) were checked on 2% agarose gels.

3.8. RT-PCR (Reverse Transcription-Polymerase Chain Reaction)

3.8.1. Reverse Transcription (RT)

cDNAs were synthesised from varying amounts of purified poly $(A)^+$ or total RNA using the SuperScript™ First-Strand Synthesis System for RT-PCR (Invitrogen).

The priming method used during this work for cDNA synthesis reaction was the use of a gene-specific primer (GSP), containing sequence information of the target cDNA.

Following use of this system, target cDNA can be directly amplified with specific primers by PCR but the sensitivity of PCR from cDNA can be increased by removal of the RNA template by digestion with RNase H after first-strand synthesis.

❖ ***Denature***

- Mix 8 µl RNA (2 µg) with 1 µl primer (GSP - 2 µM) and 1 µl dNTP mix (10mM)
- Incubate at 65ºC for 5 min., and then place on ice for at least 1 min.

❖ ***cDNA synthesis***

- Reaction mix: 2 µl 10x RT buffer
  4 µl 25 mM $MgCl_2$
  2 µl 0.1 M DTT
  1 µl RNaseOUT
  1 µl SuperScript II RT (50 U)
- Add 10 µl of reaction mix to each RNA/primer mixture (10 µl)

- Mix gently, shortly centrifuge and incubate at 42°C for 50 min.

❖ ***End reaction and remove RNA***

- Terminate the reaction at 70°C for 15 min. Chill on ice
- Add 1 µl of RNase H to each tube and incubate for 20 min. at 37°C

3.8.2. Amplification of the target cDNA (PCR)

PCRs were carried out in 50µl of final solution as protocol described below.

❖ ***PCR protocol for targets up to 4 kb***

Reaction mix: 5.0 µl	10 x PCR buffer
1.5 µl	50 mM $MgCl_2$
1.0 µl	10 mM dNTPs
1.0 µl	10 µM sense primer (for)
1.0 µl	10 µM antisense primer (rev)
0.4 µl	5 U/µl Taq DNA Polymerase
2.0 µl	cDNA (from first-strand reaction
50.0 µl	final volume with dH_2O

Cycling conditions were as follows:

First denaturation	94°C	4 min
Denaturation	94°C	1 min ⎤
Annealing	60°C	1 min 35 cycles
Elongation	72°C	1 min ⎦
Final extension	72°C	10 min*

* The PCR products to be cloned using the TOPO TA cloning kit (see 3.9) should have a final 10 minutes extension step.

The primers were chosen in the DAZ-cDNA, with the GSP (Gene Specific Primer) annealing to the 3' UTR region and the PCR primers spanning two exons boundaries to ensure the amplification of cDNA only.

3.9. TOPO® cloning

TOPO® TA Cloning provides a highly efficient one-step cloning strategy for the direct insertion of *Taq* polymerase-amplified PCR products into a plasmid vector. The plasmid vector (pCR®2.1-TOPO®) has a single 3' thymidine (T) overhang for TA Cloning and a Topoisomerase I covalently bound to the vector.

❖ ***PCR products***

- Set up a 50µl PCR reaction using *Taq* polymerase and be sure to end with a final 10 min. elongation step
- Analyse 10µl of the PCR sample by agarose gel electrophoresis
- If it is not possible to obtain a single band from the PCR, gel-purify the fragment with the QIAquick Gel Extraction kit

❖ ***TOPO® cloning reaction***

- For transformation into chemically competent *E. coli*

0.5 - 4 µl	Fresh PCR product
1 µl	salt solution
1 µl	TOPO® vector
6 µl	final volume with sterile water

- Mix gently and incubate for 1-30 min. at room temperature. For small PCR products (<1kb) the time reaction should be around 15 min.
- Place tubes on ice

❖ ***One Shot® Chemical transformation***

- Thaw One Shot® *E. coli* on ice
- Add 2µl of the TOPO® Cloning reaction to a vial of One Shot® *E. coli* and mix
- Incubate on ice for 5 min.
- Heat-shock the cells for 30 sec. at 42°C without shaking

❖ ***Recovery and plating***

- Add 250µl of SOC medium and shake the tubes at 37°C for 1 hour
- Spread 100-150µl from each transformation onto LB plates containing 50µg/ml ampicillin or kanamycin, X-Gal and IPTG
- Incubate at 37°C overnight
- Pick 10 white or light blue colonies for analysis

❖ ***Analysis of positive clones***

- Culture the 10 colonies in LB medium containing ampicillin or kanamycin, isolate plasmid DNA and analyse the plasmids by restriction analysis or by sequencing or use PCR to directly analyse positive transformants with M13 primers. For PCR use 10µl template obtained from inoculation of a single colony in 100µl TE buffer

3.10. Cell culture

3.10.1. Culture of human cell lines

Cell lines (NT2, Ep2102, Du-145 and LNCaP) grow as adherent monolayers. These cell lines were selected because they derive from testis and prostate tumour tissues. They replicate fast and are therefore easy to handle. Routinely, cells were maintained at 37°C in a humidified 5% CO_2 atmosphere. Under optimal conditions (see 2.13.2), all cells double every 24-48h. The cells were normally subcultered twice a week.

3.10.2. Trypsinisation of human cell cultures

After sucking off the old medium, cells were washed twice with PBS. Subsequently, the appropriate volume of Trypsin-EDTA solution was added to flasks and the cells were incubated at room temperature for 5 min.. The trypsinisation was stopped and cells were resuspended in complete growth media. After centrifugation (1200 rpm for 5 min.) the supernatant was removed, the pellet was resuspended in 2-5 ml fresh growth media and cell number counted.

3.10.3. Freezing and thawing of human cell lines

Cells from a healthy log-phase culture were spun down under sterile conditions for 5 min. at 1200 rpm. The cell pellet was maintained on ice, carefully resuspended in cold freezing medium (4°C) by pipetting the suspension up and down repeatedly (the cell density was at least 2×10^6 cells/ml). Aliquots of 1 ml were quickly dispensed into freezing vials. The cells were frozen slowly, first at -20°C for some hours, then overnight at -70°C. On the next day they were transferred to liquid nitrogen.

In order to thaw cells, a freezing vial was removed from liquid nitrogen and gently agitated in a 37°C water bath. The surface of the vial was rinsed with 70% ethanol and the cells were dispensed in 20 ml pre-warmed growth medium. The medium was changed once after 12 h to minimise the DMSO concentration.

3.10.4. Methylation studies on human cell lines

These studies were done to check the methylation status on the *DAZ* genes and its possible interference with its transcriptional activity (Azacytidine).

3.10.4.1. Azacytidine (AZA) treatment

Azacytidine is a potent growth inhibitor and cytotoxic agent. It acts as a demethylating agent by inhibiting DNA methyltransferase (Broday 1999). DNA methylation is an important regulatory mechanism of gene expression, gene activation and silencing.

The cells were cultivated as described on 3.10.1. After changing the medium, 5-azacytidine solution in a concentration of 1.2μg/ml was added to 5 different flasks, named A1-A5 with the numbers indicating the number of days when the cells will be harvest after treated with 5-azacytidine. An additional (A/T) flask was treated for 2 days with 5-azacytidine and subsequently treated with Trichostatin (see 3.10.4.2.).

After the azacytidine treatment, the cells were trypsinisated as described below, and the pellets were splitted into 3 different tubes and labelled as DNA, RNA and Protein. The DNA tubes were used to isolate DNA for *DAZ*-SNV/STS analyses and methylation specific PCR, the RNA tubes were used to RT-PCR protocols and the Protein tubes were stored at -80°C for posterior expression studies.

3.10.4.2. Trichostatin (TCA) treatment

In order to check if the histones acetylation plays a role on *DAZ* transcription, the cells were treated with a 300nM solution of Trichostatin A, that inhibits histone deacetylase at nanomolar concentrations, with the resultant histone hyperacetylation leading to chromatin relaxation and modulation of gene expression (Yoshida 1995). This mechanism may be involved in cell cycle progression of several cell types, inducing cell growth arrest at both G1 and G2/M phases. In some cases may induce apoptosis.

After cells (A/T flask) being treated for 1 day with 5-azacytidine, they were subsequently treated with Trichostatin A. 250 μl of a 330nM TCA solution was added to 25ml specific culture medium and the culture was incubated 7 hours at 37°C.

After this treatment, the cells were trypsinisated as described below and the pellets were collected into 1 tube labelled as RNA. The isolated RNA was used to RT-PCR protocols.

Trichostatin A is soluble at 1mg/ml in ethanol giving a clear, colourless solution.

3.10.4.3. Bisulfite treatment

DNA is modified by sodium bisulfite treatment converting unmethylated, but not methylated, cytosines to timine. Following removal of bisulfite and completion of the chemical conversion, this modified DNA is used as a template for PCR.

❖ ***Bisulfite treatment***

- Mix 2.5 μg DNA (50 μl in TE buffer) with 5.5μl 3M NaOH solution and incubate at 37°C for 15 min.
- Prepare 2.8M Na-bisulfite and 40 mM Hydroquinone. **Working solution:** mix 3.25 ml Na-bisulfite solution, 250μl hydroquinone and 150μl 10M NaOH
- Denature DNA samples at 95°C for 5 min. and place on ice
- Add 500μl of working solution to each denatured sample, add mineral oil to (5-6 drops) and incubate overnight at 55°C, in dark

❖ ***DNA purification***

To purify the DNA after the bisulfite treatment, the Wizard® DNA Clean-Up system was used with a vacuum manifold

- Add 1 ml of Wizard® DNA Clean-Up resin to 500μl treated DNA sample and mix by invertion
- Use one Wizard minicolumn for each sample. Attach the syringe barrel to the Luer-lok extension of each minicolumn
- Insert the minicolumn/syringe barrel assembly into the vacuum manifold
- Pipette the resin/DNA mix into the syringe barrel. Apply a vacuum to draw the solution through the column
- Break the vacuum to the minicolumn and add 2 ml of 80% isopropanol to the columns and re-apply the vacuum
- Dry the resin by continuing to draw a vacuum for 30 sec. after the solution has been pulled through the column
- Transfer the column to a 1.5 ml tube and centrifuge the column at 14000 rpm for 2 min.

❖ ***DNA elution***

- Transfer the column to a new 1.5 ml tube and apply 50μl of pre-warmed (80°C) TE buffer to the column
- Wait 1 min. and centrifuge the column at 14000 rpm for 20 sec. to elute the bound DNA
- Discard the minicolumn and add 5.5μl 3M NaOH
- Incubate at 37°C for 15 min. and neutralise with 55 μl 6M ammonium acetate pH 7
- DNA precipitation with 2 volumes absolut ethanol followed by wash with 70% ethanol
- Air-dry pellet, resuspend in 25μl TE buffer and store at -20°C or do methylation specific PCR (MSP)

The methylation specific PCR (MSP) is a simple method to determine the methylation status of CpG islands and it was performed using a standard PCR reaction with specific primers design to the *DAZ* promoter CpG island (DAZBis

for/rev). MSP uses the sequence differences between methylated and unmethylated alleles, which occur after sodium bisulfite treatment.

3.11. DNA sequencing

DNA sequencing was performed using an ABI PRISM 310 genetic analyser from Applied Biosystems.

❖ ***Cleaning PCR products***

- Before sequencing reaction, the PCR products were cleaned with a Microspin™ columns S-400HR from Amersham Pharmacia Biotech
- Vortex and centrifuge the column at 2900 rpm for 1 min.
- Apply the PCR product (50-100µl) to the column and centrifuge at 2900 rpm for 2 min.

❖ ***Sequencing with Big-Dye Terminator kit***

- Reaction mix: 2.5 µl Big-Dye-T mix (diluted 1:2 with halfBD)
  1.0 µl DMSO
  1.0 µl primer (2µM for or rev)
  6.0 µl PCR product+water (variable and depending on the PCR)

- Cycle programme: 94°C 30 sec.⎤
  50°C 15 sec. 25 cycles
  60°C 4 min. ⎦

❖ ***DNA precipitation***

- Add 80µl dH_2O to the sequencing sample
- Mix DNA with 10µl 3 M Na-Acetate (pH 4.6) and mix gently
- Add 250µl absolute ethanol and incubate at room temperature for 15 min.
- Centrifuge at 13000 rpm for 20 min.
- Discard supernatant, wash pellet with 250µl 70% Ethanol and mix by vortexing
- Centrifuge at 13000 rpm for 10 min.
- Air-dry pellet and resuspend in 20µl Hi-Di buffer. Store at -20°C until loading the sample

3.12. Photometric quantification for nuceic acids concentration

The nucleic acids (DNA and RNA) concentration and purity can be measured in a photometer at a wavelength of 260 nm, where the maximum absorvance for nucleic acids can be measured.

1 OD_{260} corresponds to: 50 µg/ml double strand DNA
40 µg/ml single strand DNA or RNA
20 µg/ml Oligonucleotide

The DNA concentration [μg/ml] = $OD_{260\ nm}$ × dilution factor × 50

The RNA concentration [μg/ml] = $OD_{260\ nm}$ × dilution factor × 40

By measuring the $OD_{280\ nm}$, the maximum absorvance for protein, it is possible to evaluate the purity of each sample (DNA or RNA) by the coefficient OD_{260}/OD_{280}. The value should be between 1.8 and 2.0.

Results

4.1. Mapping and structure of 4 *DAZ* gene copies in the RPCI-11-AZFc BAC contig

Earlier studies suggested that seven *DAZ* copies were clustered in distal Yq (Gläser *et al.* 1998), but a Genomic sequencing project identified only four genes in the *AZFc* region, arranged in two clusters - *DAZ*1/*DAZ*2 and *DAZ*3/*DAZ*4 with head to tail arrangement (Saxena *et al.* 2000).

The first aim of this work concerning the study of *DAZ* gene family was to start with the identification of each *DAZ* gene copy and its detailed structure in the RPCI-11-BAC sequences, using GenBank, the database of the international Human Genome Sequencing project. To identify all *DAZ*-RPCI-11-BAC clones, the GenBank database was screened *in silico* with the complete *DAZ* cDNA sequence CT351Y (GenBank accession no. AF414184). Other PACs were identified as *DAZ* clones by STS analyses in the Vogt lab (Zeisler PhD) and are listed in table IX.

Seven *DAZ*-BACs, marked in red on table VII, were gathered in two small contigs of the large AZFc contig constructed by a continuous analysis of the sequence overlap of 42 RPCI-11 BAC sequences. The AZFc contig starts with BAC 400I17 containing the STSs sY142 (DYS230, GenBank accession nr. G38345, sequence pos. 3068-3268) and sY143 (DYS231, GenBank accession no. G38347, sequence pos. 3074-3506). Both STSs have been identified as markers of the proximal AZFc border line (Simoni *et al.* 1999). It ends with the BAC 557B9, which contains STS sY167 (DYS248, GenBank accession nr. G12008, sequence pos. 98157-98132). This STS is present in the Y chromosome of all the 43 patients studied with an AZFc deletion and was therefore designated as a marker for the distal AZFc border line. The *DAZ* BAC-contig with 289L7, 140H23, 263A15 and 70G12 was mapped proximal to the *DAZ*-BAC contig with 290O3, 26D12, 539D10, having a molecular distance of approximately 1.47 Mb between them (table VII).

The proximal *DAZ*-BAC contig contains two *DAZ* gene copies (*DAZ*1 and *DAZ*2) with a 3´-5´:: 5´-3´ head-to-head orientation, starting with *DAZ*1 3´end in BAC 289L7 and ending with *DAZ* 2 3´end in BAC 70G12 (fig. 15). *DAZ* 1 contains 3 tandem repeats of 10.8 kb, including exon 2 till exon 6, besides a tandem repetitive structure of nine variants of exon 7. According to the

AZFc contig

RP11	400I17	553C13	66M18	95B23	109G18	477B5	427G18	123G1	5C5	473E1	86G22	289L7
Ac. Nr.	Ac024236	Ac007322	Ac007359	Ac023342	Ac025227	Ac007320	Ac008175	Ac016694	Ac010080	Ac016911	Ac006366	Ac010088
Size (bp)	19622	184215	100627	69354	60810	187894	205237	19704	156956	30824	166460	199458
Obs.	sY142/sY143	pY6H65			sY139	Transposon sY139	pY6H65		50f2/C	Start56267 of 5C5	Dys237 sY150	Start 74683 of 86G22
GENES		RBMY1	PRY1		RBMY1	insert after 16909	RBMY1	PRY1?		End 47152 of 86G22	BPY2/TTY4	DAZ1

RP11	140H23	263A15	70G12	39P20	221K4	506M9	589P14	823D8	978G18	1067I16	307L15	533E2
Ac. Nr.	Ac053490	Ac007039	Ac006983	Ac009947	Ac016707	Ac016752	Ac025246	Ac073649	Ac073893	Ac068601	Ac023274	Ac012005
Size (bp)	103432	17107	180345	149230	17537	166439	29964	120951	64450	75756	137682	189557
Obs.	Start28292 of 289L7		1-27620 rep. DAZ	pY6H65	St140004 of 39P20/	DYS240	Star 30619 of 506M9			sY110, 109	End 32878 of 533E23	
GENES	DAZ1/2	DAZ2	DAZ2						TTY3	CDY1	CSPG4LY	GOLGA2LY

RP11	424J12	160O2	535I13	214M24	290O3	26D12	539D10	363G6	566H16	102O5	497C14	251M8
Ac. Nr.	Ac013465	Ac016698	Ac010153	Ac025735	Ac010089	Ac006982	Ac006338	Ac016728	Ac006386	Ac006328	Ac007562	Ac010682
Size (bp)	7664	154286	104147	85472	103356	2147	186838	152052	164229	177769	137077	145383
Obs.		Sy150	Star 102493 of 160O2					sY150		GOLGA2LY		
GENES			BPY2/TTY4		DAZ 3	DAZ3/4	DAZ 4	BPY2/TTY4		CSPG4LY	CDY1	TTY3

RP11	79J10	100J21	245K4	270H4	487K20	557B9
Ac. Nr.	Ac007273	Ac017005	Ac007965	Ac006991	Ac024067	Ac013734
Size (bp)	149794	161879	182083	173242	108346	180019
Obs.	Ulincomp.			pY6H65		sY167

Table VII – *AZFc* RPCI-11 BAC contig. The two *DAZ* contigs and the corresponding BAC clones are marked in red. Other BAC clones are marked according to the palindrome where each BAC belongs (e.g. BACs in blue belong to blue palindrome (b1-b4) on AZFc). Observations concerning STS content and BACs overlapping are added for a better understanding of this contig. The genes, except *DAZ*, are labelled in light blue. STS identified as borderlines (sY142 and sY167) are also indicated.

nomenclature of Yen *et al.* (1997) they were designated as *A, B, C, D, E, F, X, Y, Z* with the Y variant as the most prominent consensus sequence (fig. 16).

Eight variants were found in *DAZ*1 and organised in the order: 5´-*B-C-D-E-F-E-X-Y-Z*-3´ (fig. 15). Since identical repetitive exon 7 structures linked to a triplicate *DAZ* exon 2-6 domain have been found in testicular *DAZ* cDNAs (Saxena *et al.* 2000), it was assumed that the *DAZ*1 gene is transcribed in human testis.

The *DAZ*2 gene structure is mapped to three overlapping BAC sequences. Typical of *DAZ*2 is the presence of the *A* variant and absence of the *B* variant in the repetitive exon 7 structure, besides 7 tandem *Y* variants (fig. 15). *DAZ2* transcripts with this exon 7 variant structure have been found in testicular *DAZ* cDNA sequences (Yen *et al.* 1997; Saxena *et al.* 2000). This indicates that the *DAZ*2 gene is also transcribed in human testis. However, no *DAZ*2 testis transcripts containing a duplicated *E-F* variant motif were reported, nor an exon 7*D* variant flanked proximal by exon 7*X* or distal by exon 7*E*, as it was predicted by the genomic *DAZ2* sequence in BAC263A15 (fig. 15).

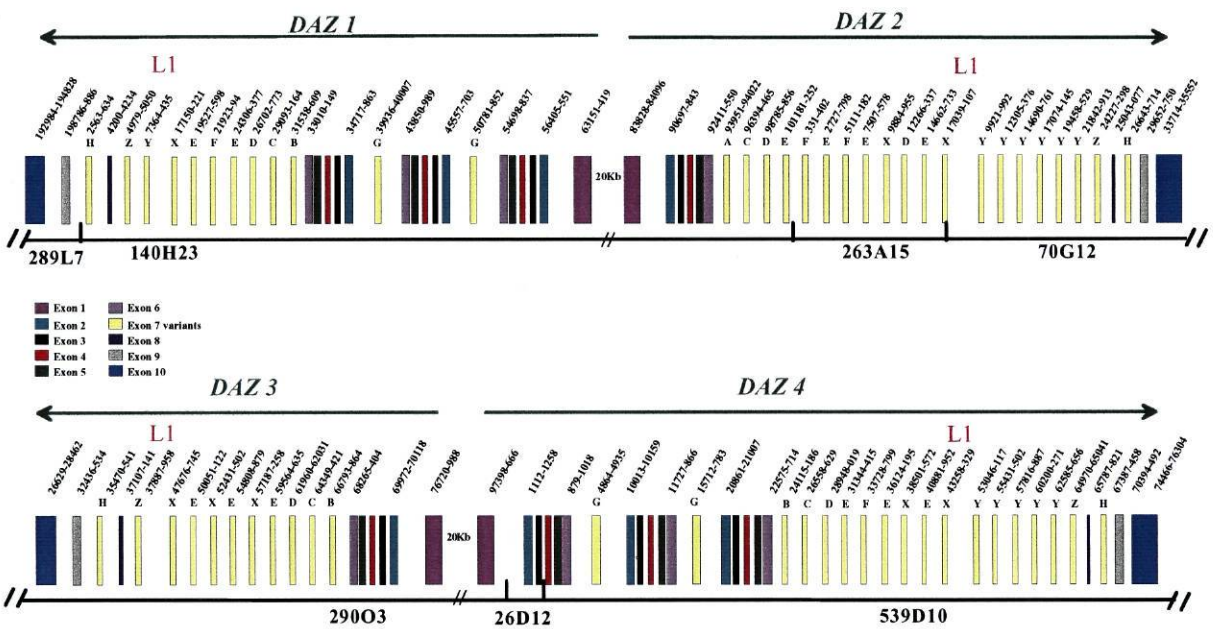


Figure 15 – Exon structure of DAZ-BAC contigs. The BAC numbers are indicated above each structure.

Such *DAZ2* transcripts might therefore be rare or do not exist and an artefact of the sequence assembly programmes used for the construction of the genomic BAC 263A15 sequence has produced this exon 7 variant composition. The structure of the *DAZ2*-BAC sequence of the CTB library (BAC352E14) also contains an *A* and 7 tandem *Y* variants in the repetitive exon 7 structure but the duplicated *E-F* motif found in the RPCI-11 BAC263A15 sequence was absent (Saxena *et al.* 2000).

In the distal region of the AZFc contig (table VII), 1.47 Mb distal to the 3´end of *DAZ2*, the sequence of BAC 290O3 contains the third *DAZ* gene copy, followed by the overlapping BACs 26D12 and 539D10 with the fourth *DAZ* gene copy (fig. 15). Although both have a similar 3´-5´:: 5´-3´ head-to-head orientation as found for *DAZ*1 and *DAZ*2 copies, substantial differences were found in their repetitive exon structures.

```
7Y:   GCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATATAATTATCAG
7A:   --T-A-T-----------CA-------GG--------T--------G--------T----------------
7B:   --T-ACT-----------CA-------GG--------T--------G--------T----------------
7C:   -A--A----A-------CGA----G-------------------------------------------------
7D:   C-------------------A----------------TG---------------------------------
7E:   -------------------A------------A---G----------------C--------C---------
7F:   C-------------------------------------TG---------------------------------
7H:   A---A----A---------A-------GG--------------G-G--------------G-----------
7G:   ---CACA--A--C-C----A----G--G----C-G--T----A-------T---------------------
7X:   -------------------A-------G---------------------------------C---------
7Y:   -------------------------------------------------------------------------
7Z:   -------------------A----G--G-----------------------C-A------C-----C---
```

Figure 16 – DAZ exon 7 variants. In red the 7Y variant defined as consensus sequence; in blue, the 7 variants never found in human testis cDNA.

Typical for *DAZ3* in the RPCI-11-BAC 290O3 sequence is the absence of the *Y* variant and a triplication of the 7 *E-X* motif in the repetitive exon 7 structure (fig. 15). The *DAZ4* gene structure in the RPCI-11 BAC 539D10 is characterised by three copies of a 10.8 kb repeat unit. In the RPCI-11-*DAZ4* gene, 15 exon 7 variants were found with the following composition: 5´-*B-C-D-E-F-E-X-E-X-Y-Y-Y-Y-Y-Z*-3´. Although testicular *DAZ* cDNAs with this exon 7 variant typical structure (i.e. *B* variant linked to 5 tandem *Y* variants) were described by Yen *et al.* (1997), complete *DAZ*4-cDNAs with 15 exon 7 copies, linked to a triplicate 10.8 kb repeat have not yet been isolated.

Two additional exon 7 variants with normal splicing signals were found in intron 6 of the first and second 10.8 kb repeat (exon 7*G* variant) of *DAZ* 1 and *DAZ* 4

and in the intron 8 sequence (exon 7*H* variant) on all genomic *DAZ* copies gene structure (fig. 15). Because no *DAZ* cDNA sequence containing one or both of these *DAZ* exon 7 variants has been found in human testis, *DAZ*1 transcripts spliced with these variants must therefore be rare or do not exist in testis tissue.

4.2. Establishment of a set of DAZ-SNVs (Single Nucleotide Variants)/ STSs (Sequence Tagged Sites) to distinguish each *DAZ* gene copy

DAZ gene locus restriction maps and sequence analyses have shown that the genomic structures of the 4 *DAZ* genes present in the AZFc region are more than 99% homologous to each other (Saxena *et al.* 1996, 2000; J. Zeisler, S. Fernandes, unpublished results). For the identification of *DAZ* gene copy specific deletion events, the search for *DAZ* copy specific markers along the complete gene sequences was therefore focused on a systematic identification of small sequence variants, including a specific restriction site (*DAZ*-Single Nucleotide Variants *DAZ*-SNVs). Typing the SNVs involves co-amplification of two or more similar loci followed by restriction enzyme digestion to reveal the absence (A-allele unrestricted) or presence (B-allele restricted) of a given restriction site. Beside this I searched for copy-specific *DAZ*-STSs in the RPCI-11-*DAZ*-BAC sequences. 6 *DAZ*-SNVs (I-VI) spanning the 5´end till the 3´end of *DAZ* gene sequences as well as 2 *DAZ* gene copy specific STSs (*DAZ*-RRM3, Y-*DAZ3*) were established (table VIII).

SNV II was first identified by Agulnik *et al.* (1998). SNV III and SNV V were identified by Saxena *et al.* (2000). SNV I is mapped 90 nucleotides upstream of the putative transcription initiation site of the *DAZ* genes with a polymorphic *Fsp*I restriction site (B allele) only present in *DAZ4*. SNV IV was found 1159 nucleotides downstream of *DAZ* exon 7*H* in the intron 8 sequence with a polymorphic *Alu*I restriction site absent in *DAZ2*. SNV VI was found 480 nucleotides downstream of *DAZ* exon 10, where due to a 4bp deletion (CCTG) at the 3´end of *DAZ4*, an *Afl*II site was created. An extension of this deletion to 8 nucleotides (*T*CCTG*TAT*) was identified at the 3´end of the *DAZ3* gene and used for the development of a *DAZ3* specific STS (Y-*DAZ3*).

DAZ-SNVs	DAZ gene region	GeneBank accession nr	Restriction enzyme	DAZ gene copy	SNV allele	Fragments (bp)
I	5´ UTR-exon1	G73167	Fsp I	1,2,3 4	A B	709 398+311
II	exon3-exon4	G73166	Mbo I	1 2,3,4	A B	182 122+60
III (sY586)	intron 7A/B	G63907	Taq I	2 1,3,4	A B	301 184+117
IV	intron 7H	G73168	Alu I	2 1,3,4	A B	660 398+262
V (sY587)	intron 9	G63908	Dra I	3,4 1,2	A B	195+49 122+73+49
VI	3´ DAZ	G73169	Afl III	1,2,3 4	A B	431 248+183
DAZ-STSs	**DAZ gene region**	**GeneBank accession nr**	**BAC clone(s)**	**Present in DAZ copy**	**Tanneal °C**	**PCR product (bp)**
DAZ-RRM3	DAZ1+ 4 intron 6b*	G73171	BAC140H23 BAC539D10	1+4	66	473
Y-DAZ3	3´ DAZ	G73170	BAC290O03	3	66	260
DAZ1+4 (sY152)	DAZ1+4 intron 6a*	G12039	BAC140H23 BAC539D10	1+4	66	125
SY14	SRY	G38356	-	Control	66	470

Table VIII - *DAZ*-SNVs and STSs, with different alleles in *DAZ1-DAZ4* gene copies. STS/SNV alleles in bold letters and numbers mark their specificity for one *DAZ* gene copy; * Intron 6a and 6b are proximal (a) and distal (b) to exon 7G present in 10.8 kb repeat in *DAZ*1 and *DAZ*4 (fig.15).

A second STS (*DAZ*-RRM3) was established in the fused intron1-intron 6 structure specific for *DAZ1* and *DAZ4*. This marker is in the same intronic region as sY152 declared earlier to be a marker for the *DAZ1* and *DAZ4* genes (Saxena *et al.* 2000). However, we also found the sY152 primer pair sequences in the RP11CI-*DAZ2* and -*DAZ3* gene sequences (after exon 7*A* and 7*B*, respectively) with only 3 internal mismatches in the forward primer (fig 17). Therefore we often obtained false positive amplification products in our PCR experiments and developed the *DAZ*-RRM3 STS as a more stringent alternative for the identification of the *DAZ1* and *DAZ4* gene copies using our PCR multiplex conditions. All novel *DAZ*-SNVs and STSs including the optimal PCR conditions were deposited in the GenBank under the accession numbers given in table VIII.

sY152 in BAC 539D10 (DAZ4)

```
2581 cttttaaact agtgaactgt acctaaaatt taagaaaaca cttagaatta gtgtaatgaa
2641 gacctctgta gtatgtagaa gtgatgaaat aatattttga caggagggta cttagcagta
2701 acttttctgc agaacagttt ctgagatttg gtgttcctgt cttggtttca gcatgaattt
2761 tgttatcttt gctgtcaaag agctgaaaca tggcagactg tctttcgtga attttgtaga
2821 gatacagagt gaaataaaag ctttacccaa ttcttagagc gcagaattcc aattgtgttt
```

sY152 in BAC 140H23 (DAZ2)

```
94141 gaaatcactt agaattagtg taatgaggac ctctgtttta tttagaagtg atgaaataag
94201 attttgacag gagggtactt agcattaact tttccgtaga acaatttctg agatttggtg
94261 ttcccttctt tgtttcagca tgtattttgt tatctttgct gtcaaagagc tgaaacatgc
94321 agactgtctt tcctgaatct tgtagagata cagagtgaaa taaaagcttt accgaattct
94381 tagagcacag aatttcagtt gtatttttat tttagcttgc tgcttcatga tagcagttct
```

Figure 17 - sY152 primers (in pink) in *DAZ*4 copy showing 100% homology and in *DAZ*2 copy showing 3 internal mismatches (in green) in the forward primer and 1 mismatch in reverse primer.

4.2.1. *DAZ* SNV/STS analysis in *DAZ*-PACs and *DAZ*-BACs

The establishment of the novel *DAZ*-SNVs and STSs was firstly used for the identification of the *DAZ* gene copies present in the *DAZ*-PACs and BACs isolated before. The results are summarised on table IX with allele A meaning non-digested, allele B digested and A+B non-digested and digested.

PAC/cosmids	SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	sY152	RRM3	DAZ 3	DAZ gene
G19633C	-	A	B	B	B	A	-	-	-	1 without RRM
G19660C	A+B	B	B	B	A	A	+	-	-	*DAZ*4
K07760F	A	A	B	B	B	A	+	++	-	*DAZ*1
D061100C	A	A	B	B	A	A	-	+	-	not defined
O131180D	A	A	B	B	B	A	+	+	+	*DAZ*1
K07760E	A	A	B	B	B	A	+	+	-	*DAZ*1
cos63C9	-	A	B	B	B	A	-	-	-	*DAZ*1part
BAC clones	**SNV I**	**SNV II**	**SNV III**	**SNV IV**	**SNV V**	**SNV VI**	**sY152**	**RRM3**	**DAZ 3**	**DAZ gene**
290O3	A+B	B	B	B	A	A	-	-	+	*DAZ*3/4part
140H23	A	A	A+B	B	B	A	+	+	-	*DAZ*1/2
539D10	A	B	B	B	A	B	+	++	+w	*DAZ*4
226N11	A	B	A	A	B	A	-	-	-	*DAZ*2
26D12	A+B	B	B	B	A	A+B	+	+	+	*DAZ*3/4
263A15	A	A+B	A	A	B	A	+	+	-	*DAZ*2/1
148I14	A	A	B	B	B	A	+	+	-	*DAZ*1
CTB352E14	A	B	A	A	B	A	-	+w	-	*DAZ*2
CTA132B16	A	B	B	B	A	A	-	-	+	*DAZ*3
CTB546E5	A	A	B	B	A	A	+	++	-	*DAZ*4
K1020B	A	A	B	B	B	A	+	+	-	*DAZ*1

Table IX – DAZ SNV/STS analyses in *DAZ*-PACs and *DAZ*-BACs.

Based on these set of SNVs and STSs, specific *DAZ* deletion haplotypes (*DAZ*-del-hap) were established, each one indicative for the deletion of a specific set of *DAZ* gene copies (table X). The identification of one of them would predict the deletion of the related *DAZ* gene copy. Partial *DAZ* gene deletions would be indicated by the occurrence of different *DAZ* deletion haplotypes.

DAZ gene deletions	SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	DAZ-RRM3	Y-DAZ3	DAZ del haplo
DAZ1	A+B	**B**	A+B	A+B	A+B	A+B	+	+	**1**
DAZ2	A+B	A+B	**B**	**B**	A+B	A+B	+	+	**2**
DAZ3	A+B	A+B	A+B	A+B	A+B	A+B	+	-	**3**
DAZ4	**A**	A+B	A+B	A+B	A+B	**A**	+	+	**4**
DAZ1+ DAZ2	A+B	**B**	**B**	**B**	**A**	A+B	+	+	**5**
DAZ1+ DAZ3	A+B	**B**	A+B	A+B	A+B	A+B	+	-	**6**
DAZ1+ DAZ4	**A**	**B**	A+B	A+B	A+B	**A**	-	+	**7**
DAZ2+ DAZ3	A+B	A+B	**B**	**B**	A+B	A+B	+	-	**8**
DAZ2+ DAZ4	**A**	A+B	**B**	**B**	A+B	**A**	+	+	**9**
DAZ3+ DAZ4	**A**	A+B	A+B	A+B	**B**	**A**	+	-	**10**
DAZ1+ DAZ2+ DAZ3	**B**	**B**	**B**	**B**	**A**	**B**	+	-	**11**
DAZ1+ DAZ2+ DAZ4	**A**	**B**	**B**	**B**	**A**	**A**	-	+	**12**
DAZ2+ DAZ3+ DAZ4	**A**	**A**	**B**	**B**	**B**	**A**	+	-	**13**

Table X - *DAZ* deletion haplotypes (*DAZ*-del-hap).

4.2.2. *DAZ* SNV/STS analysis in fertile men (FM)

On the Y chromosome of men with normal fertility, the presence of 4 *DAZ* gene copies would be indicated by the presence of A and B alleles of each of the 6 *DAZ*-SNVs and the presence of the STSs *DAZ*-RRM3 and Y-*DAZ3*. The deletion of a specific *DAZ* gene copy is recognised by the absence of specific SNVs A or B alleles associated with this *DAZ* gene copy (table X) and the deletion pattern of the *DAZ*-RRM3 and Y-*DAZ3* STSs. Absence of all *DAZ* markers would indicate the absence of the complete *DAZ* gene family on the Y chromosome.

One hundred and seven fertile men (FM samples) were used as a control population and the results obtained are shown on table XI.

Surprisingly, 3 DAZ-del-hap could be found in this control group, indicating the existence of some *DAZ* partial deletions even in fertile men.

SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	DAZ-RRM3	Y-DAZ3	FM samples
A+B	A+B	A+B	A+B	A+B	A+B	+	+	50
DAZ-del-hap 2 and derivatives								
A+B	A+B	**B**	**B**	A+B	**A**	+	+	25
							-	1
A+B	A+B	**B**	A+B	A+B	A+B	+	+	1
A+B	A+B	**B**	A+B	A+B	**A**	+	+	4
A+B	A+B	A+B	**B**	A+B	A+B	+	+	1
A+B	A+B	A+B	**B**	A+B	**A**	+	+	5
							-	3
DAZ-del-hap 4 and derivatives								
A	A+B	A+B	A+B	A+B	**A**	+	-	2
A	A+B	A+B	A+B	A+B	A+B	+	+	6
A+B	A+B	A+B	A+B	A+B	**A**	+	+	3
DAZ-del-hap 9 and derivatives								
A	A+B	**B**	**B**	A+B	**A**	+	+	2
							-	1
A	A+B	**B**	A+B	A+B	**A**	+	+	1
A	A+B	A+B	**B**	A+B	**A**	+	+	1
							-	1

Table XI - DAZ-del-hap analyses in 107 men with proven fertility.

DAZ-del-hap 2, indicative for a deletion of *DAZ* 2 was found in 26 FM samples and is associated with the absence of the B allele for SNV VI, suggesting an additional deletion in the distal part of *DAZ*4. Derivatives of *DAZ*-del-hap 2 (absence of only the A allele of SNV III or SNV IV) were also identified in this control group. Similar pattern was also observed for *DAZ*-del hap 4 (indicative for deletion of *DAZ*4) and *DAZ*-del hap 9 (indicative for deletion of *DAZ*2 and *DAZ*4).

The results suggest that these *DAZ* deletions are not associated with an oligozoospermic phenotype and could be seen like familial variants, inherited from father to son.

4.2.3. *DAZ* SNV/STS analysis in oligozoospermic patients (OZ)

We selected for the analysis of *DAZ* gene copy specific deletions 63 infertile men with a clinical diagnosis of idiopathic oligozoospermia. These men had a normal karyotype and no Yq11 microdeletions.

DAZ-del-hap 2 and derivatives were found in 20 OZ samples; derivatives of DAZ-del-hap 4 were found in 5 samples and 1 sample showed the *DAZ*-del-hap 9. Beside these 3 *DAZ*-del haplogroups, found already in fertile men group, we

found in 5 OZ samples *DAZ*-del hap 5, indicative for a deletion of *DAZ*1 and *DAZ*2.

SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	DAZ-RRM3	Y-DAZ3	OZ samples
A+B	A+B	A+B	A+B	A+B	A+B	+	+	32
DAZ-del-hap 2 and derivatives								
A+B	A+B	**B**	**B**	A+B	**A**	+	+	8
A+B	A+B	**B**	A+B	A+B	**A**	+	+	2
A+B	A+B	A+B	**B**	A+B	A+B	+	+	1
A+B	A+B	A+B	**B**	A+B	**A**	+	+	4
							-	5
DAZ-del-hap 4 and derivatives								
A	A+B	A+B	A+B	A+B	A+B	+	+	2
A+B	A+B	A+B	A+B	A+B	**A**	+	+	3
DAZ-del-hap 9 and derivatives								
A	A+B	**B**	A+B	A+B	**A**	+	+	1
DAZ-del-hap 5 and derivatives								
A+B	**B**	**B**	**B**	**A**	A+B	+	+	2
A+B	**B**	**B**	**B**	**A**	A+B	-	+	1
B	**B**	**B**	**B**	**A**	**A**	+	+	2

Letters in bold indicate the presence of deletions

Table XII - DAZ-del-hap analyses in 63 oligozoospermic men.

Two of them showed also absence of the A allele for SNV I and of the B allele for SNV VI (fig. 18), indicating additional partial deletion on the 5´region of the *DAZ*3 and in the 3´region of the *DAZ*4 gene, respectively.

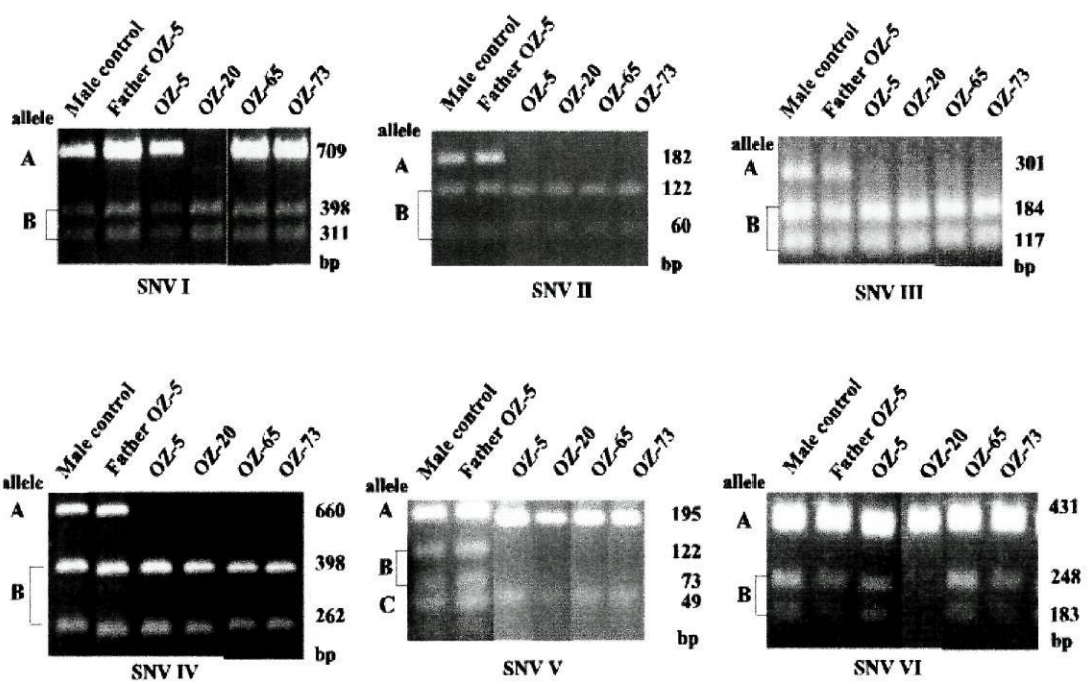


Figure 18 - *DAZ*-SNVs analysis of DNA samples with *DAZ*-del-hap 5.

Some results obtained for *DAZ*-SNVs from DNA samples with *DAZ*-del-hap 5 (defined on table X) are shown in figure 18. On table XII are summarised the results obtained for the studied oligozoospermic population.

The *DAZ*-SNV/STS analyses suggests that the B alleles of SNV I and VI in *DAZ*4 are polymorphic. The derivatives of *DAZ*-del-hap 2, 4 and 9 were found in FM and OZ samples. We conclude that most probably, familial gene conversion events occurred between the *DAZ* different copies, especially within the *DAZ*2 and *DAZ*4 genes. Because gene conversion events prevent the identification of a specific *DAZ* copy deletion by only SNV/STS analyses, an additional analysis by DYS1 blot experiments was developed.

4.3. DNA blot analysis of DYS1 locus

The Y chromosome specific p49f /*Taq* I polymorphism of the DYS1 locus was the first one described (Ngo 1986) and is now known to be part of the *DAZ* locus (fig 7).

The p49f probe was sub-cloned from cosmid 49 (Bishop *et al.* 1984). It is an *Eco*R I fragment of 2.847 bp, starting in intron 3 of the third *DAZ1* RRM domain (BAC140H23 sequence pos.: 34027) and ending in *DAZ1* intron 7B (BAC140H23 sequence pos.: 31181; GenBank accession no. AF414183). After *Taq* I digestions, 18 restriction fragments could be detected (table XIII) and after *Eco*RV digestions, 5 restriction fragments cross-hybridised with the p49f probe in fertile men population (table XIII).

After the establishment of the *DAZ* sequence structure in AZFc on a BAC contig (fig. 15), we were able to map the different genomic DYS1 fragments to each *DAZ* gene copy (table XIII). There are identically sized fragments arising from more than one copy but some of them (marked in red on table XIII) could be identified to be associated with a specific *DAZ* gene copy.

Taq I	DAZ1	DAZ2	DAZ3	DAZ4	%homology
A4			19612		81
A3					
A2				14853	<80%
A1		10100			81
B	10094				<80%
C0					
D1		7152			81
D2					
E	5854/7			5857(2)	97/83
F1			5008		83
G	4782	4781		4783	81
H	4020	4020		4021	83
I		3112			99/81
J	2975	2976	2975	2975	83
K (DAZL)					83
L (DAZL)					86
M	2076	2076	2075	2075	81
N	1992(3)	1992	1992	1992(3)	99/97
O	1877		1876	1876	99
P		1419(3)		1419(2)	80
Q	1237		1236	1236	81
R	966	966		966(2)	80
EcoRV	**DAZ1**	**DAZ2**	**DAZ3**	**DAZ4**	**% homology**
	20321	20320	20321	20323	99/86
	10843/5				99/98
				7325/7	97/83
	5169	5170	5169	5169	83
	2385-91	2382-96	2390	2385-90	81

Table XIII – *Taq* I and *Eco*R V restriction fragments (in kb) co-hybridising with p49f probe.

Two DYS1-*Eco*RV fragments, 10845 and 10843 bp, could be associated with *DAZ*1 and two DYS1-*Eco*RV fragments, 7327 and 7325 bp with *DAZ*4. Similarly, the 3112 bp DYS1-*Taq* I fragment could be associated with *DAZ*2 and the 19612 bp DYS1-*Taq* I fragment with *DAZ*3. Therefore it should be possible to identify the deletion of a single *DAZ* gene copy by deletion of the associated DYS1-*Eco*RV and *Taq* I fragments. This prediction was controlled with DNA blots from gels of *DAZ*-BACs after restrictions with *Taq* I and *Eco*R V. The results are shown on figure 19.

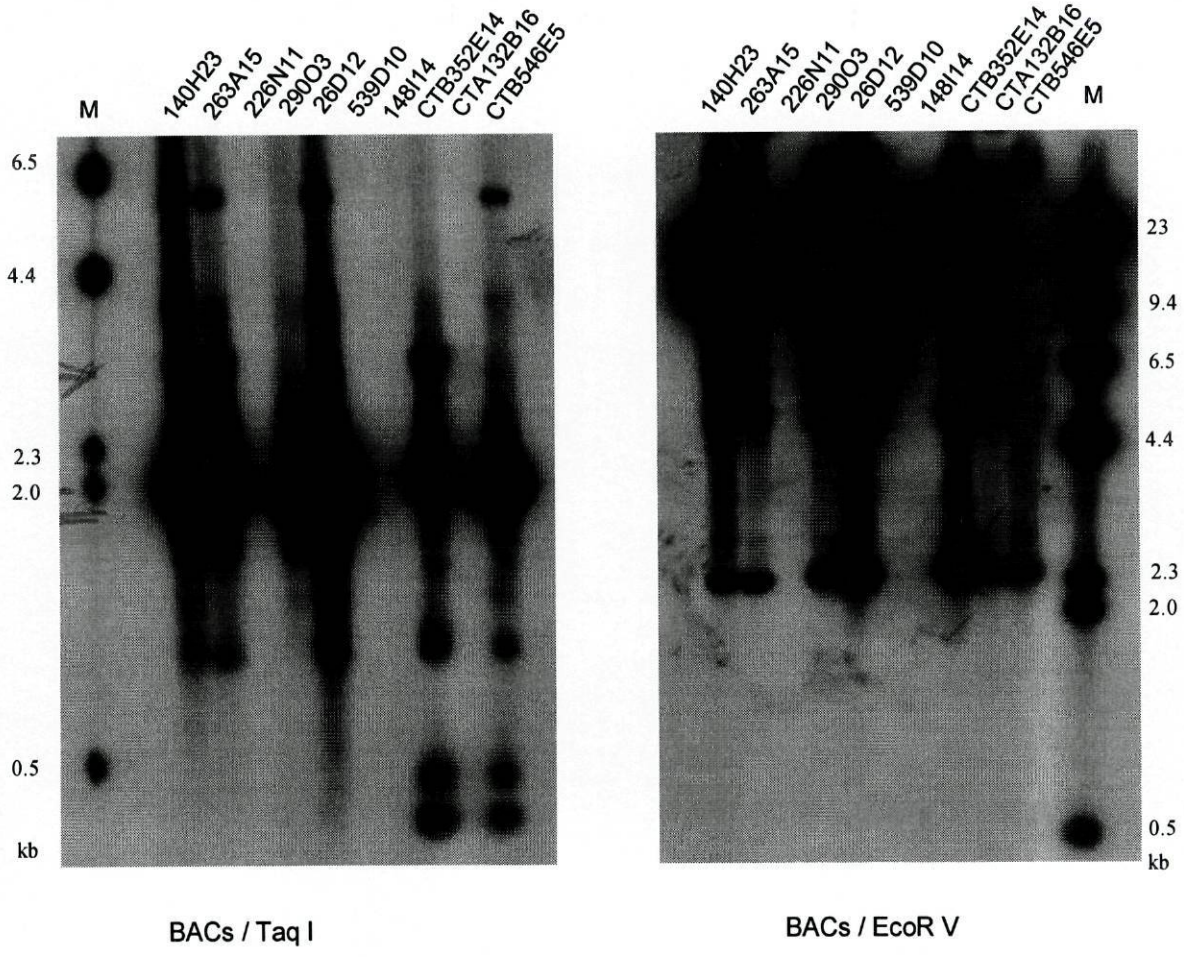


Figure 19 – Taq I and *Eco*R V restrictions using BACs DNA as template and p49f probe for hybridisation.

Due to the low quality of the hybridisation results obtained, the 226N11, 539D10 and 148I14 BACs were excluded from this analysis. The results obtained (fig. 19) confirmed some of the predictions for the *DAZ* content, described on table II: the presence of the 10.8 kb *Eco*R V fragment in BACs 140H23 and CTB546E5 indicative for *DAZ*1 copy; the presence of the 3.1 kb *Taq* I fragment for BACs 140H23, 263A15 and CTB352E14 indicative of *DAZ*2; the presence of the 7.3 kb *Eco*R V fragment in BAC 26D12 indicative of *DAZ*4 gene copy; the 19.6 kb *Taq* I fragment indicative for *DAZ*3 is not visible on BAC 290O3, probably due to the weakness of this sample. In BAC CTB546E5 the *Eco*R V fragment indicative of *DAZ*4 gene copy has a higher size, probably indicative of the different *DAZ* structure on Y chromosome from the male donor of the CTB library.

Although the BAC-blots experiments did not allow the identification of single *DAZ* gene copies, it also does not contradict the *DAZ* SNVs/STSs analyses. So, this system was used in cases where enough amount of DNA (~20ug) was available.

Some results from samples with *DAZ*-del-hap 5 are shown in figure 20. The deletion of the *DAZ*1 and *DAZ*2 genes in the samples with *DAZ*-del-hap 5 was confirmed by absence of the *DAZ*1 10.8 kb-*Eco*R V and *DAZ2* 3.1 kb-*Taq* I fragments of the DYS1 locus group. OZ-5 and his father displayed a familial deletion of *DAZ*4, detected through the absence of the *DAZ*4 specific 7.3 kb *Eco*R V fragment in both DNA samples.

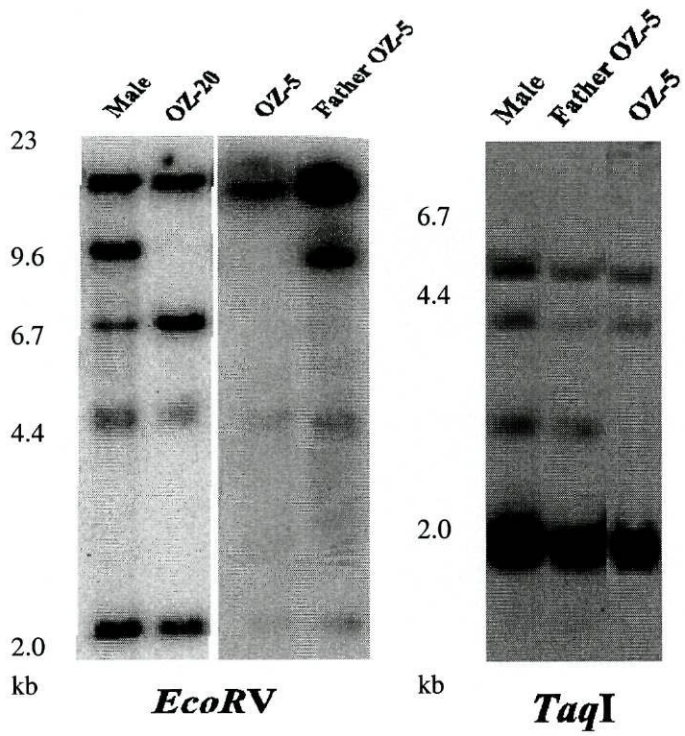


Figure 20 – Taq I and *Eco*R V restrictions using genomic DNA (*DAZ*-del-hap 5 samples) as template and p49f probe for hybridisation. Absence of 10.8kb-*Eco*R V fragment on OZ20 and OZ5, indicative for deletion of *DAZ*1 copy; absence of 3.1kb-*Taq* I fragment on OZ5, indicative for *DAZ*2 deletion; absence of 7.3kb-*Eco*R V fragment on OZ5 and his father, indicative of a familial deletion of *DAZ*4. Pictures used with permission of Vogt & Hirschmann.

However, we could confirm that the *DAZ*1/*DAZ*2 gene deletion in the OZ-5 family is restricted to the patient OZ-5, i.e., is a *de novo* deletion.

Deletions of the *DAZ*3 and *DAZ*4 genes in the Y chromosomes of OZ-74 (data not shown) and OZ-75 (*DAZ*-del-hap 10) were confirmed by absence of the

*DAZ*3 19.6 kb-*Taq* I fragment and *DAZ*4 7.3 kb-*Eco*R V fragment (Fig. 21). The variable length of the A band is explained on table XIII with *DAZ*3 being responsible for 19.6 kb fragment, *DAZ*4 for 14.8 kb and *DAZ*2 for 10.1 kb *Taq* I fragment. The prediction that the *DAZ*3 and *DAZ*4 gene deletions in OZ-74 and OZ-75 are not associated with their testicular pathology was also confirmed because the fertile father of OZ-5 has the same *DAZ* deletion (Fig. 20). A DNA sample from father of OZ-74 was not available.

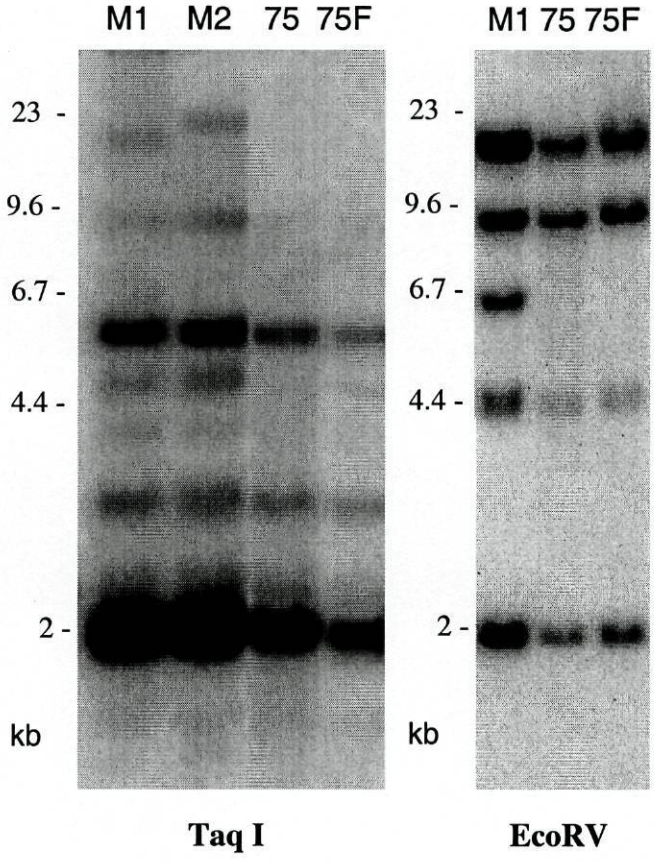


Figure 21 – Taq I and *Eco*R V restrictions using genomic DNA (*DAZ*-del-hap 10 samples) as template and p49f probe for hybridisation. Absence of 19.6kb-*Taq* I fragment and absence of 7.3kb-*Eco*R V fragment on sample 75 and his father indicative, respectively, for a DAZ3 and *DAZ*4 familial deletion. Pictures used with permission of Vogt & Hirschmann.

4.3.2. *Mlu* I blots

A *Mlu*I restriction map, specific for each *DAZ* gene copy, was reported by Saxena *et al.* (2000). By *in silico* analyses a specific Mlu I fragment of 13kb was found for *DAZ*1 and *DAZ*4, 21kb for *DAZ*2 and >75kb for *DAZ*3, with a 633bp fragment present in all *DAZ* copies serving as a positive control for the Mlu I restriction digest. On figure 22 (left) are presented some examples of the

experiments performed with *DAZ*-BACs. The 13 kb fragment present in BACs CTA148I14, 140H23 and CTB546E5 is indicative of *DAZ*1 copy for the first two and of *DAZ*4 copy for the last one and the 21 kb fragment is indicative of *DAZ*2 copy. However, in none of the genomic samples (fig. 22 right) these fragments were detected, indicating no Mlu I restriction. It was assumed that this fact is related with the methylated state of these DNAs, presenting most of the Mlu I sites methylated and, therefore, not available for Mlu I restriction activity.

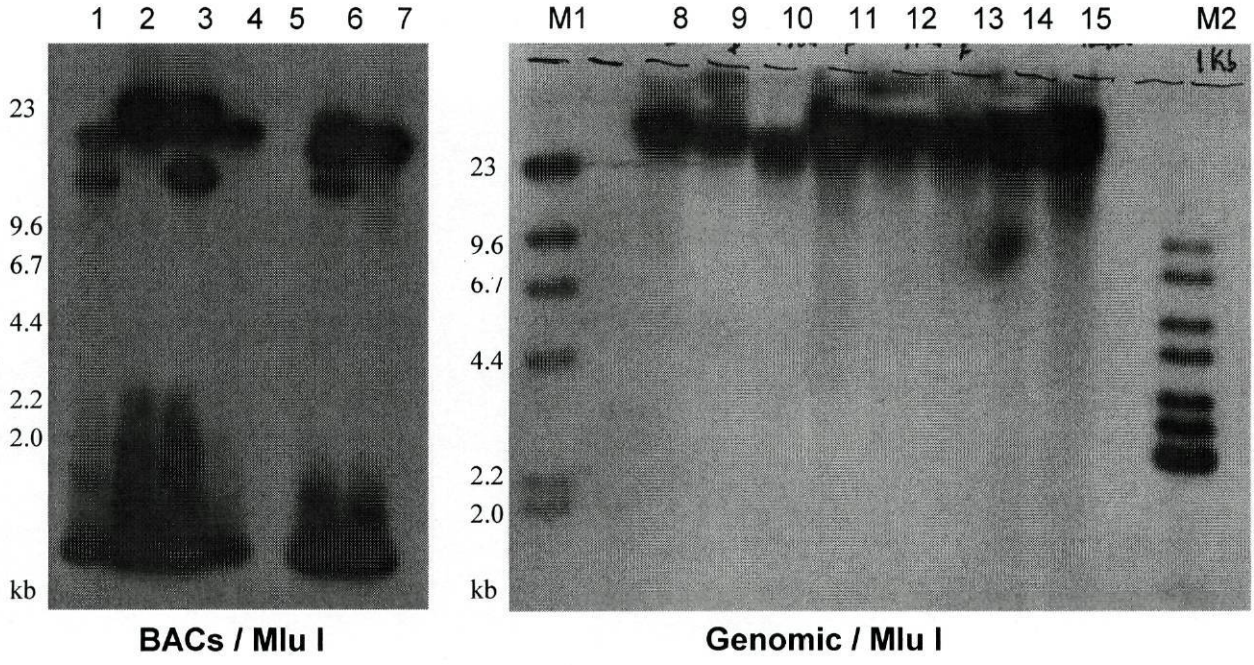


Figure 22 – BACs (non-methylated DNA) and genomic DNA blots after Mlu I restriction using DAZ1.1 as genomic probe. The samples are as following: 1-148I14; 2-CTB352E14; 3-CTB546E5; 4-CTA132B16; 5-290O3; 6-140H23; 7-K1020; 8,9-fertile male control; 10-OZ5; 11-OZ5 father; 12-OZ5 AZA1w; 13-OZ5 father AZA1w; 14-OZ5 AZA2w; 15-OZ5 father AZA2w; M1-λ/HindIII; M2-1kb ladder.

Samples 12 to 15 were treated with 5-azacytidine (1 and 2 weeks) in order to demethylate the genomic DNA but no relevant differences were observed. Therefore, it was not possible to check the methylation state of these samples after this treatment.

4.4. SNVs analysis of AZFc genes and amplicons

During the *DAZ* gene copy analysis was developed, the complete *AZFc* sequence has been published by Kuroda-Kawaguchi *et al.* (2001).

This structure is composed of large repeated sequence blocks called amplicons. There are six distinct families of amplicons in AZFc region (fig. 23). The amplicon unit range in length from 115kb (grey amplicon) to 678kb (yellow). The turquoise, grey and yellow amplicons occur twice, whereas the green amplicon occurs three times and the blue and red amplicons occur four times each. Together the six amplicon families account for 93% of the 4.5Mb sequence defined as AZFc. This palindromic complex (fig. 23) contains 11 families of transcription units, most of them with exclusive testis expression: 15 genes representing 5 families previously reported to have protein coding potential – *RBMY*1, *PRY*, *BPY*2, *DAZ* and *CDY*1; six new families of transcription units – *CSPG4LY*, *GOLGA2LY*, *TTY*3, *TTY*4, *TTY*5 and *TTY*6 with the last four presenting non coding transcripts.

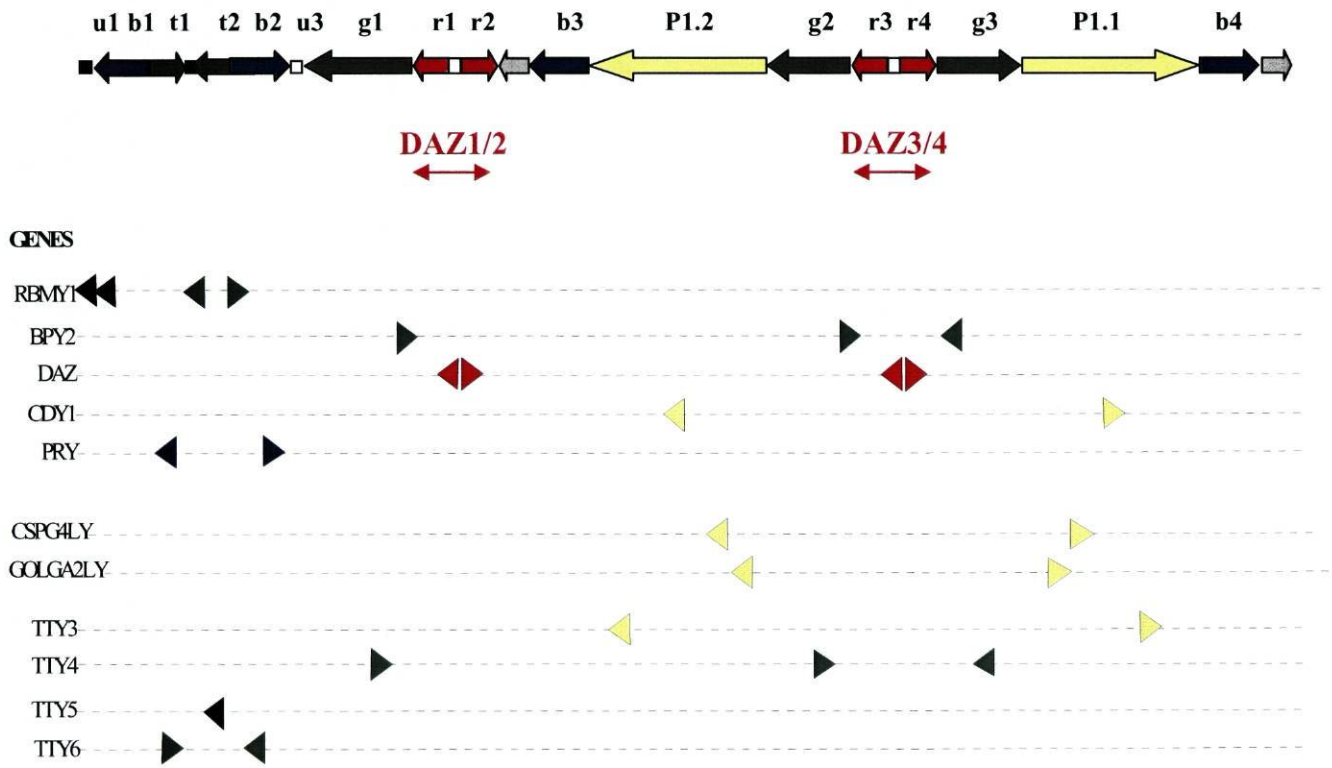


Figure 23 – *AZFc* palindromic structure and listed below, the AZFc genes with a functional transcription unit. Their 5'-3' polarities correspond to the arrows directions, coloured according to their corresponding amplicons.

It was therefore possible to investigate the variation in all gene families in the AZFc region, outside the *DAZ* genes. The sequence data of Kuroda-Kawaguchi

et al. (2001) was used and single nucleotide variants were searched, as have been found in the *DAZ* gene locus, within exonic and intronic structure of other six gene families: BPY2 (Basic Protein Y2), CDY1 (Chromo-Domain Y1), CSPG4LY (Chondroitin sulfate proteoglycan 4-like Y), GOLGA2LY (Golgi antigen 2-like Y), TTY3 (Testis transcript Y 3) and TTY4 (Testis transcript Y 4). SNVs were found for BPY2 and TTY4 mapped in the green amplicons (g1, g2, g3) and GOLGA2LY in the yellow amplicons (P1.1, P1.2). It was not possible to find suitable variants in the CDY1, CSPG4LY and TTY3 genes located in the yellow amplicons (fig. 23). Therefore, an AZFc-Yellow marker located outside the genes but able to distinguish the yellow P1.1 amplicon from the P1.2 was developed. The summary of this AZFc SNV set is shown on table XIV and presented graphically on figure 24.

SNVs	AZFc site	GeneBank accession nr	Restriction enzyme	AZFc amplicon	SNV allele	Restricted fragments (bp)
Yellow SNV I	P1.1/P1.2	G73351	Dde I	P1.1 P1.2	A B	467 284+183
GOLY SNV I	GOLGA2LY genes (2 copies)	BV012731	Hha I	P1.1 P1.2	A B	531 289+242
BPY2 SNV I	BPY2 genes (3 copies)	BV012732	EcoRV	g2, g3 g1	A B	470 289+181
TTY4 SNV I	TTY4 genes (3 copies)	BV012733	Hae III	g3 g1, g2	A B	541 323+218
STSs	**AZFc site**	**GeneBank accession nr**	**BAC clone(s)**	**BAC seq. (position)**	**Tanneal °C**	**PCR product (bp)**
SY1192	u3	G67166	BAC 5C5 AC010080	52109-52362	64	255

Table XIV - AZFc SNVs and STS with different alleles in palindromes and gene copies.

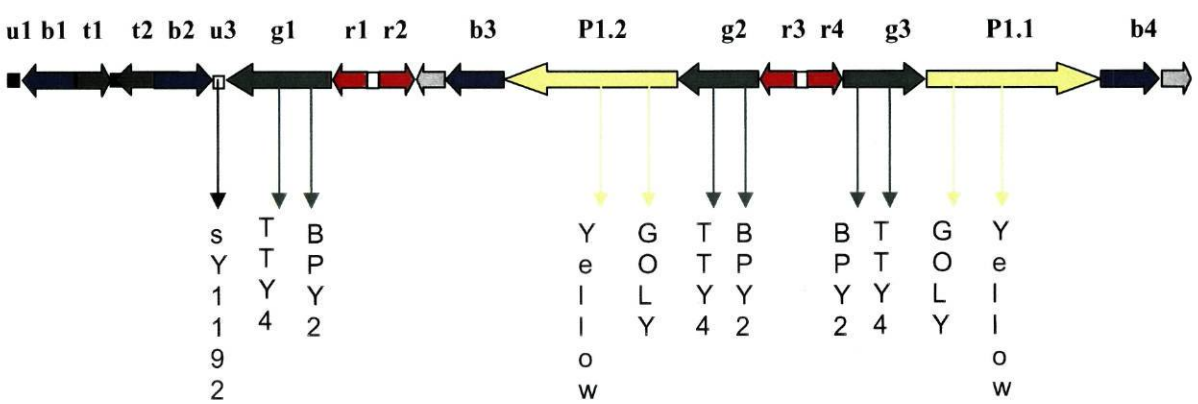


Figure 24 – *AZFc* palindromic structure and the localisation of the AZFc-SNVs developed.

Thirty one samples from fertile men (In1-In31), with variable *DAZ*-SNVs were analysed with this *AZFc* marker set (table XIV). Variations in SNVs that do not affect fertility, due to deletion, gene conversion or other causes, will behave like neutral markers and can be placed on the phylogeny to reveal the number of such events and some information about their likely time and place of origin.

These samples were also typed by S. Paracchini, with a large set of Y-chromosomal binary markers, defining 116 haplogroups, to establish their location in the phylogeny, according to Underhill *et al.* (2000) and Paracchini *et al.* (2002). Only 10 of the 116 haplogroups were present in these samples, as would be expected from their predominant origin in Europe (table XV). From these results, thirteen different "AZFc haplotypes" could be defined in these samples, 12 of which were associated with the loss of one or a small number of variants in DAZ- and AZFc-SNVs set (fig. 25). Among these, five belonging to a common haplogroup (haplogroup N), lacked SNVs from a large segment of the distal part of the AZFc containing the *DAZ*3 and *DAZ*4 genes.

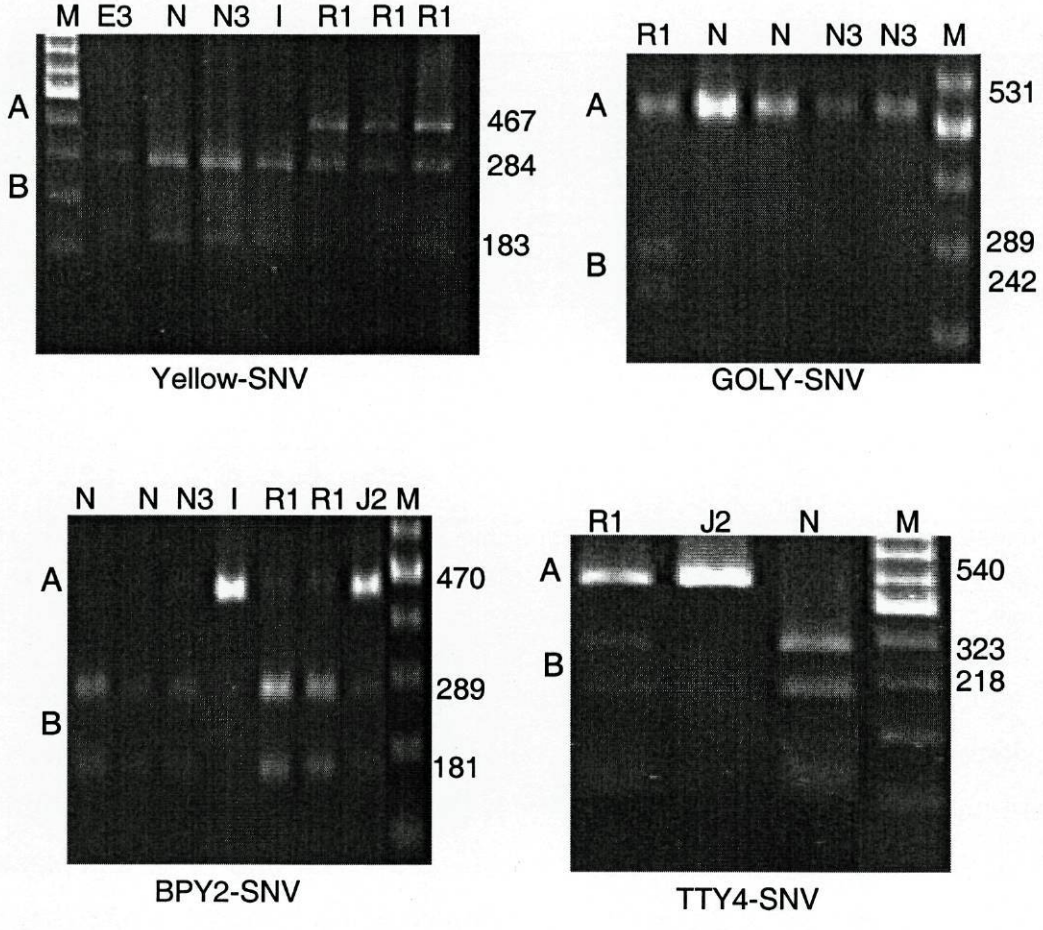


Figure 25 – AZFc-SNVs analysis of DNA samples from different Y-haplogroups.

DNA-sample	DAZ-SNV I	DAZ-SNV II	DAZ-SNV III	DAZ-SNV IV	DAZ-SNV V	DAZ-SNV VI	Y-DAZ3	50f2/C	sY1192	TTY4-SNV I	BPY2-SNV I	GOLY-SNV I	AZFc-P1-SNV I	AZFc del ht	Y-haplogroup	YCC-code
In1	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In2	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A	A+B	B	2	108	R1a1*
In3	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In4	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In5	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A	A+B	B	2	108	R1a1*
In6	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In7	A+B	A+B	B	B	A+B	A	+	+	+	A+B	A	A+B	B	3	52	I*
In8	A+B	A+B	B	B	A+B	A	+	+	+	A+B	A	A+B	B	3	52	I*
In9	A+B	A+B	B	B	A+B	A	+	+	+	A+B	A	A+B	B	3	52	I*
In10	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A+B	B	B	4	58	J2*
In11	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A	A+B	B	5	35	E3b1*
In12	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A+B	A+B	B	6	56	J2f*
In13	A+B	A+B	B	B	A+B	A	+	+	+	A	A	A+B	B	7	58	J2*
In14	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A	A+B	B	5	32	E3b3a*
In15	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	8	52	I*
In16	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	8	35	E3b1*
In17	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	8	32	E3b3a*
In18	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	8	35	E3b1*
In19	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	8	35	E3b1*
In20	A	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	9	104	R1*
In21	A	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	9	104	R1*
In22	A+B	A+B	A+B	A+B	A+B	A	+	+	+	A+B	A+B	A+B	B	10	115	Q3*
In23	A	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	B	11	108	R1a1*
In24	**A**	**A+B**	**A+B**	**A+B**	**A+B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**12**	**92**	**N3***
In25	**A**	**A+B**	**A+B**	**A+B**	**A+B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**12**	**92**	**N3***
In26	A	A+B	A+B	B	A+B	A	+	+	+	A+B	A+B	A+B	B	13	76	O3*
In27	A	A+B	A+B	B	A+B	A	+	+	+	A+B	A	A+B	B	14	52	I*
In28	A	A+B	B	A+B	A+B	A	+	+	+	A+B	A+B	A	B	15	58	J2*
In29	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***
In30	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***
In31	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***

* The Y-haplotyping was performed by S. Paracchini and C. Tyler-Smith.

Table XV – Summary of AZFc-SNVs/STSs analyses and Y-haplogroup typing in 31 DNA samples.

To confirm the relationship between the *DAZ*3/*DAZ*4 deletion and the observed phylogeny, 15 individuals known to belong to the Y-haplogroup N (In32-In46) were analysed for their "*DAZ* and AZFc haplotype". The results are summarised on table XVI. Beside small variations in *DAZ*-SNVs II and IV (2 samples) and one sample with sY1192 positive, all samples lost u3 and g2-*DAZ*3-*DAZ*4-g3-P1.1 markers. It was, therefore, identified a common Y lineage, haplogroup N, related with a large deletion in this part of the AZFc region.

DNA-sample	DAZ-SNV I	DAZ-SNV II	DAZ-SNV III	DAZ-SNV IV	DAZ-SNV V	DAZ-SNV VI	Y-DAZ3	50f2/C	sY1192	TTY4-SNV I	BPY2-SNV I	GOLY-SNV I	AZFc-P1-SNV I	AZFc del ht	Y-haplogroup	YCC-code
In32	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**+**	**B**	**B**	**A**	**B**	**16**	**87**	**N*(xN3)**
In33	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**87**	**N*(xN3)**
In34	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**87**	**N*(xN3)**
In35	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**87**	**N*(xN3)**
In36	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**87**	**N*(xN3)**
In37	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***
In38	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***
In39	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***
In40	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***
In41	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**87**	**N*(xN3)**
In42	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**87**	**N*(xN3)**

* The Y-haplotyping was performed by S. Paracchini and C. Tyler-Smith.

Table XVI - Summary of AZFc-SNVs/STSs analyses in 15 DNA samples from Y-haplogroup N.

4.5. *DAZ* promoter structure

The first approach to understand the regulatory mechanism controlling testis-specific *DAZ* genes expression was the molecular analysis of the *DAZ* upstream region for putative promoter elements. For this purpose, several *in silico* analyses were performed using different promoter prediction programmes, listed on table XVII. The promoter is usually defined in terms of the sequences that react with RNA polymerase as well as a unique set of transcription factors which probably influence the transcription initiation.

Promoter programmes	Internet addresses	Reference
Promoter Inspector	http:/genomatix.gsf.de	Quandt *et al.* 1995
ProScan	http:/bimas.dcrt.nih.gov/cgi-bin/molbio/proscan	Prestridge 1995
Promoter 2.0	www.cbs.dtu.dk/services/promoter	Knudser 1999
FunSiteP	http:/transfac.gbf.de	Kondrakhin *et al.* 1995
Promoter Predictions	www.fruitfly.org/cgi-bin/seq_tools/promoter	Fickett 1997

Promoter databases	Internet addresses	Reference
Eukaryotic Promotor Database	http://www.epd.isb-sib.ch	Périer *et al.* 1998
Transcription Regulatory Regions Database	http://www.bionet.nsc.ru/trrd	IC&G SB RAS 1997

Table XVII - Promoter prediction programmes and databases.

The results for *DAZ*1 intron 1 are schematically shown on figure 26 and the most informative data obtained is summarised on table XVIII. The results pointed for a *DAZ* gene promoter domain at *DAZ* region from exon 1 to exon 2 (intron 1).

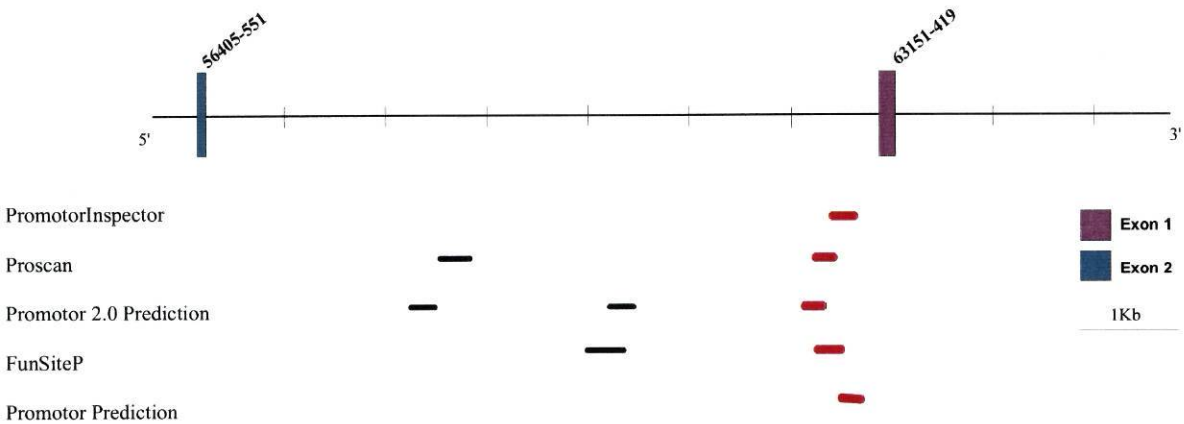


Figure 26 – Schematic representation of *DAZ*1 promoter predictions in BAC140H23, using different *in silico* tools.

DAZ gene	BAC pos.	Prom. Inspector	Pro Scan	Prom2.0	FunSite P	Prom prediction
*DAZ*1	140H23 (56000-63420)	62689-62912	62340-62680	Score>1 62500	62600-62700	62968-63018
*DAZ*2	140H23 (84061-90850)	84334-84557	84568-84808	Score>1 88461	84562-84619	_
*DAZ*3	290O3 (70971-76990)	76260-76483	76000-76250	_	_	_
*DAZ*4	290O3 (97381-103356)	97902-98129	98138-98388	_	_	_

Table XVIII – *DAZ* promoter predictions with the exact positions found in each *DAZ*-BAC.

In eukaryotes, DNA methylation affects cytosine residues and it is specific for CpG sequences. The CpG island found in the *DAZ* genes, located upstream to the translation initiation site (ATG), are shown on figure 27.

DAZ promotor

CTCCCATCCAACCCCTTATATCTCAAACTTGACAGAAAGGAGTCCAAGTTTCTGGCCTGTTGATGTCTCC

CTCCGTTTTCCACAGACTTGCTACTATTTATTTCTGCAGGTACTTTAAAAGATTTTGGGGTGGGTGTGAC

GCTAACAGTGATTGTGCTAAACCTGTGCATTCCGGAAGCCATTTTGAAACTACAGCCCAGACACATTATC

ATTTTTTAAAACTATACCATATATTTCAGTATGCATACTTATTTCAAGGGTGTTTTGGAGGCGGTGCCTG

CCTGGAACCCAGGCACTGGCCAGCACACACGCACGGCATACCCTGCGTACTCTACGCACCCACAGCGTGG

GGGACGGCCGTCATCCCCCGCAGGCCTCCACCCTCCAGGGGGCGGGTGTGCTGCCACCCATTGGCTCAGG

CCGATACCACGCGCCCCGATACCCGGCACAGGAGCCACCTCCCAGAGCCCCGCAGTCCATGCCTCAGTCG

GCCTGCGCTCCTCAGCCTGGCGGTTCTACCTCCGAGGGTTCGCCCGCCCTTGGTTTTCCTTACACCTTAG

CCTTTGGCTCCTTTGACCACTCGAAGCCCCACAGCGTGTTCCAGCGGCACTTCACCAGCAGACCCAGAAGT

GGTGGGTGAAACACTGCCTCTGTTCCTCCTTGAGCCTGTCGGGAGCTGCTGCCTGCCACCACCATGGTGA

GTTGAAGGAAAGACCTGAGGGGCAAGGCCTGTGGGGTCTTCACCGGTGGGTGAAACCAGGCTCCGCCTTA

TCCTCGGTTTTCCTGTTTCGGGGCCACCCCGCTGAATGCATTAGGAGCCGTGGCCCGCAGGGTGGGGTGA

Figure 27 – CpG island in *DAZ* promoter region found with CpG island searcher programme (http://ccnt.hsc.usc.edu/cpgisl). Legend: Exon 1, GC boxes, ATG as the start codon.

4.6. Transcriptional regulation of the *DAZ* locus by methylation?

The results obtained from the Mlu I blots with the genomic DNA samples (fig 22) are indicative for a methylation of the *DAZ* gene locus. It has been therefore analysed whether this methylation is also influencing the transcriptional activity of the *DAZ* genes.

To study the effects of methylation in *DAZ* gene expression in vitro, 2 testicular and 2 prostate cell lines (table XIX) were used, taking the advantage of the testicular origin of 2 of them and the data of Lau & Zhang (2000) showing expression in prostate. They will also provide large amount of material to be studied (DNA and RNA) and allow treatments with chemicals to induce demethylation and acetylation of the nuclear chromatin.

4.6.1. Cell lines characterisation

The cell lines characterisation studies started with cytogenetic analysis in order to check their karyotypes and their Y-chromosome integrity. The karyotypes (table XIX) were quite complex and to clarify the results, M-FISH studies were done in all cell lines. These results are presented on Appendix A.

Testicular cancer cell lines	Sex chromosome constitution
NT2	62XY
Ep2102	53XXY
Prostate cell lines	
Du-145	61XY,t(Y ;20)
LNCaP	74XYY

Table XIX – Sex chromosome constitution from 2 testicular and 2 prostate cell lines.

The Y chromosome had complete structure in both testicular cancer cell lines. Du-145 cell line presents a translocation between Y chromosome and chromosome 20 (t(Y;20)) and LNCaP had a Y-chromosomal duplication.

DAZ-SNV and *AZF*-SNV analyses were then performed to analyse the *DAZ* locus and AZFc structural integrity in all cell lines. The results obtained for *DAZ*-SNV are shown on figure 28 and for *AZFc*-SNV on figure 29. The *AZF*-SNV pattern obtained for LNCap suggests that this cell line belongs to the AZFc deletion haplotype 8, as it presents the same AZFc-SNVs pattern found for samples In15 to In19 (see table XV).

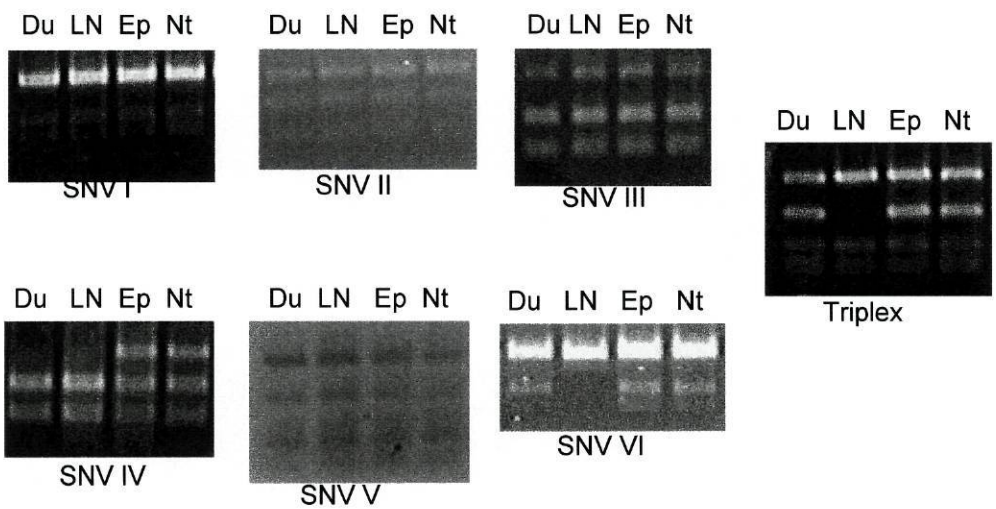


Figure 28 – *DAZ*-SNVs analysis on cell lines. Absence of SNV VI B allele and Y-*DAZ*3 STS in LNCaP.

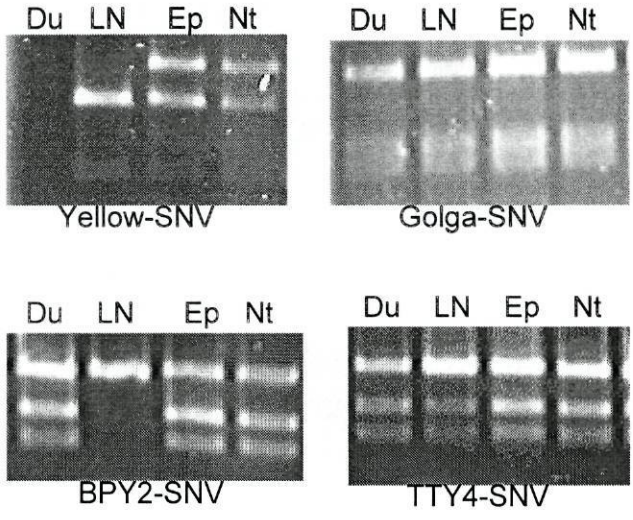


Figure 29 – *AZFc*-SNVs analysis on cell lines. Absence of Yellow-SNV A allele and B allele of BPY2-SNV in LNCaP.

The results obtained with both SNV analyses were indicative of some molecular rearrangements on *AZFc* of Y chromosome in LNCaP cell line, not only on the *DAZ* locus but specially on the *AZFc* region. In order to clarify this point, Southern blot analysis was done with the same system previously used for BACs and genomic samples (*Eco*R V blot). The results, shown on figure 30, demonstrated the *DAZ* locus molecular integrity of the four cell lines as they show the same pattern obtained for the male control.

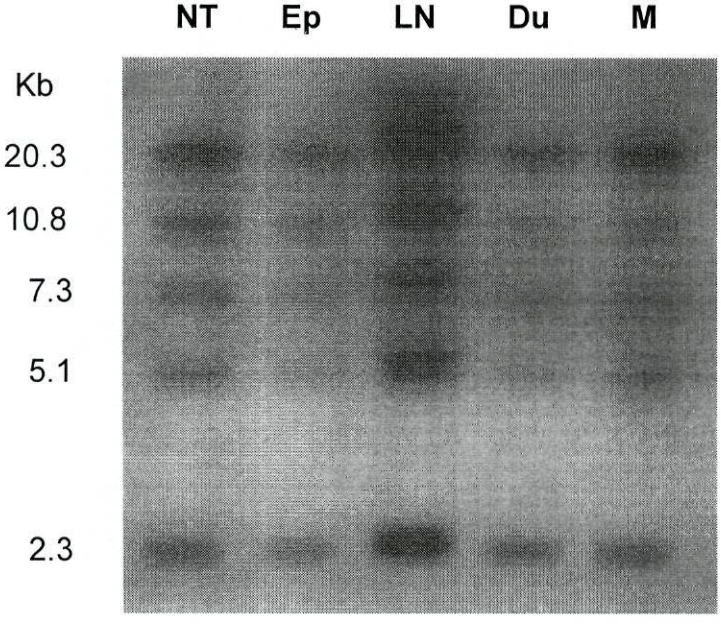


Figure 30 – *Eco*RV restrictions using cell lines DNA as template and p49f probe for hybridisation. NT - NT2; Ep - Ep2102; LN - LNCaP; Du - Du-145; M – Male control.

4.6.2. *DAZ* transcription studies

4.6.2.1. RT-PCR on cell lines

In order to study transcription of the *DAZ* genes in each cell line, specific primers were designed. The reverse transcription (RT) step was always performed with a gene specific primer covering part of the *DAZ* exons 9 and 10 (cDNADAZ primer, table I). The primers used in the RT-PCR experiences and their specificity are displayed on figure 31. The blue coloured ones, named *DAZ*-short (table I), indicate a general positive *DAZ* transcription and do not distinguish among the different *DAZ* gene copies. The lilac coloured ones, named long (DAZ1/4, table I), identify specific transcription of *DAZ*1 and *DAZ*4 copies; the orange coloured ones, named 7A (table I), identify specific *DAZ*2 transcripts. *DAZL* specific primers (355ex2-3for and 355ex8-9rev, table I), covering the exon 2/3 boundary till exon 8/9 boundary and *DBX* primers covering the cDNA from exon 3/4 till exon 6/7 were used as positive controls for all RT-PCR experiments.

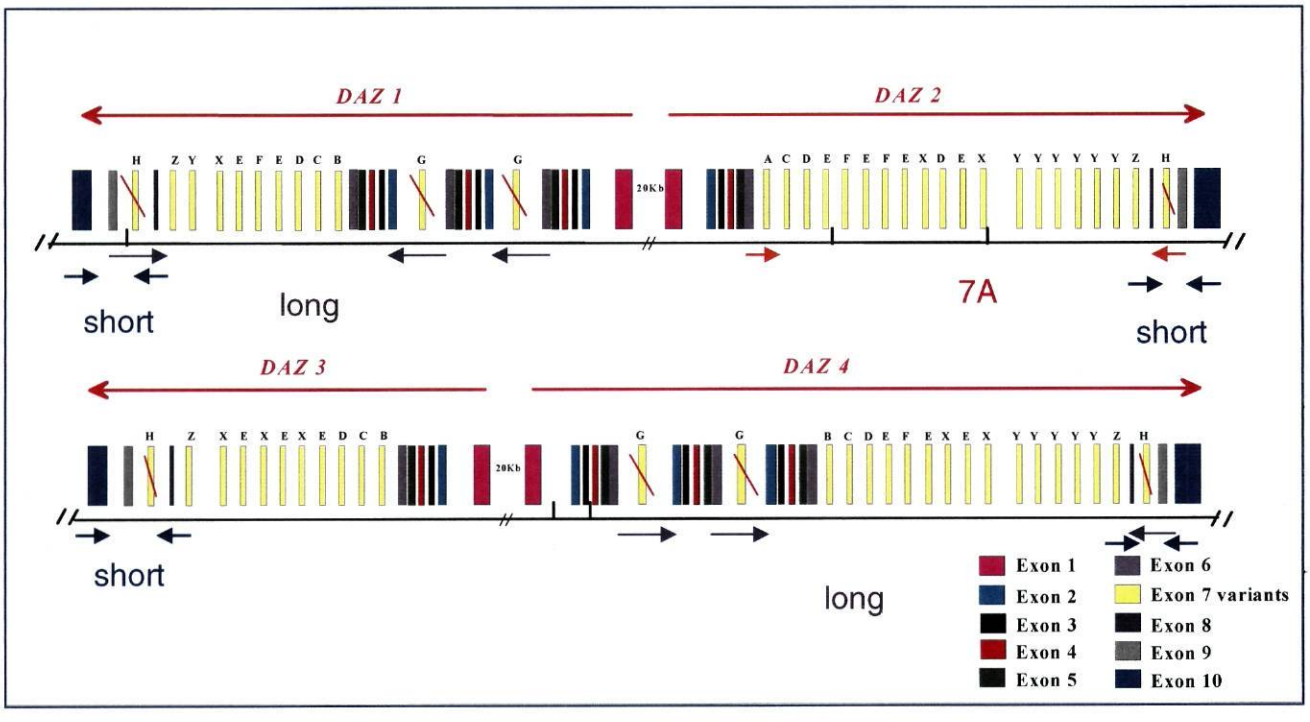


Figure 31 – Localisation of DAZ primers used for transcription studies. Because testis *DAZ* transcripts containing exons 7G and 7H were never found, they are crossed.

The RT-PCR results are displayed on figure 32. Although long transcripts, indicative of *DAZ*1 and *DAZ*4 transcription, were not found in NT2 and Du-145 cell lines, two transcripts with different lengths (1.2 and 1.6 kb) were obtained for LNCaP. As shown on fig. 31 (lilac primers) they correspond to transcripts containing one or two RRM domains, respectively. Short transcripts indicative for transcription of all *DAZ* genes were not found on Du-145 cell line. The transcript estimated length is 254bp. The existence of 2 transcripts for NT2 and Ep2102 could only be explain by the existence of different transcripts containing small exons, already described as pseudoexons (ψ8) (Saxena *et al.* 1996, fig. 5). Female and male controls do not display any transcription showing the primers specificity for cDNA. On human testis sample the DBX transcription was very weak.

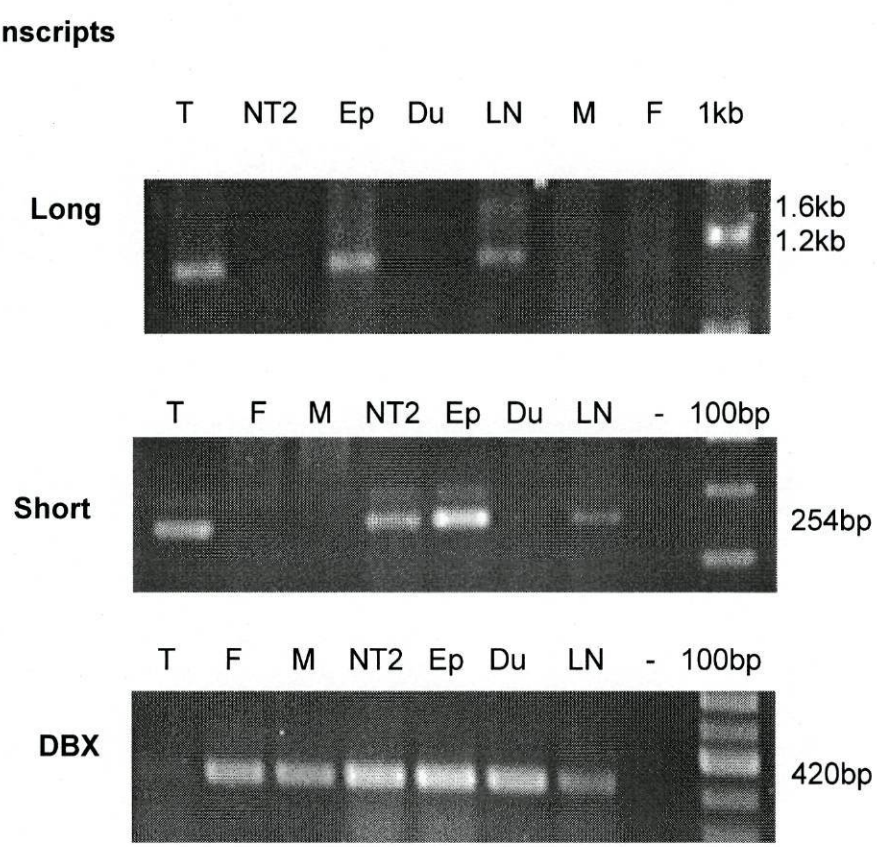


Figure 32 – DAZ transcripts on cell lines. T- human testis; F – female control; M – Male control; 100bp ladder used as a molecular weight marker.

Specific primers were also designed to check the existence of testis transcripts containing the exons (7G and 7H). The primers used were selected for annealing to exon 6 and exon 2 boundaries (primer 6-2), exon 6 and exon 7G

(primer 6-G) and exon 7G and exon 2 (primer G-2). They could produce RT-PCR products only from transcription of *DAZ* copies containing more than 1 RRM domain (*e.g. DAZ*1 and *DAZ*4). The results obtained are shown on figure 33. The Ep2102 cell line displayed two transcripts with 412 and 907bp corresponding to both RRM domains, indicating that the *DAZ* copies transcripts had 2 RRM domains.

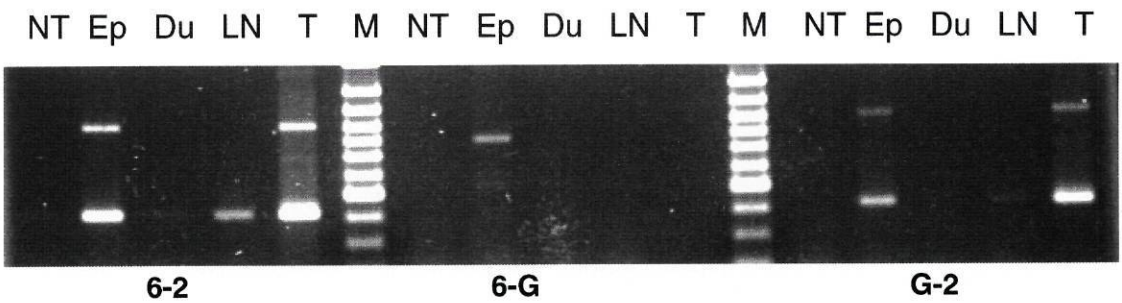


Figure 33 – DAZ transcripts on cell lines. NT - NT2; Ep - Ep2102; LN - LNCaP; Du - Du-145; T-human testis; M – 100bp ladder; No results were obtained when using primers to check the transcription of exon 7H.

The different *DAZ* transcripts obtained (fig. 32 and 33) were extracted from the agarose gel, cloned and sequenced in order to identify their origin for a distinct *DAZ* gene copy, by analyses of the number of RRM domains present in each transcript and by the presence of different 7 exon variants composition. Because repetitive sequences, like the ones from the 7 exon variants are quite difficult to get, good sequences could be only obtained from the testis transcripts. On table XX are shown the different compositions of 7 exon variants found in different transcripts.

Clone name	Long transcripts	*DAZ* gene
Testis1.4	Ex6-2,3,4,5,6...Ex7y,7y,7y,7z,8,9	*DAZ*4
Testis1.5	Ex6-2,3,4,5,6,7b,7c..Ex7x,7e,7y,7y,7y,7y,7z,8,9	*DAZ*4
Testis1.6	Ex6-2,3,4,5,6,7b,7c,7d..Ex7e,7x,7y,7z,8,9	*DAZ*1
Clone name	**Short transcripts**	
Testis 7A1	Ex7a,7c,7d...Ex8,9,10	*DAZ*2
Testis2.4	Ex7a,7c,7d,7e,7f,7e	*DAZ*2
Testis2.1	...Ex7e,7x,7e,7x,7z,8,9	*DAZ*3
Testis1.2	...Ex7d,7e,7f,7e,7x,7y,7z,8,9	*DAZ*1

Table XX – Sequencing results from cDNA *DAZ* clones.

The analysis of all sequences obtained for each transcript confirmed the transcription of all *DAZ* gene copies, at least in testis. Among long transcripts, *DAZ*4 ones could be distinguish from those from *DAZ*1 by the number of 7Y variants present (*DAZ*1 was described as containing only one 7Y while *DAZ*4 should contain 6 of the exon 7Y variant). *DAZ*2 transcripts were found quite often, suggesting that these transcripts should be produced earlier than the others. In all clones sequenced, it was not possible to find the 7H variant, between exon 8 and exon 9, as predicted by *in silico* analysis. This suggests that the transcripts containing this variant do not exist or are only produced in a low level. Extrapolation from the results obtained to the testis to the cell lines transcripts has been allowed as they presented the same size on the gel.

4.6.3. *DAZ* gene transcription after demethylation of genomic DNA

The transcriptional activity of the *DAZ* genes in the 4 cell lines was analysed, after treatment with 5-azacytidine for different time points (1 to 5 days).The results are shown on figures 34-37.

The *DAZ*7B forward primer differs from 7A (shown on fig. 31) only in the first nucleotide and it was designed in this way to be specific for the exon 7B variant, present in all but *DAZ*2 copy.

Additionally the Trichostatin A treatment was used to check if the histones acetylation plays a role on *DAZ* transcription. The results are shown on the right side of each picture, on figures 34-37, being referred as A/T, as the cell lines were treated firstly (2 days) with 5-azacytidine and then with Trichostatin A.

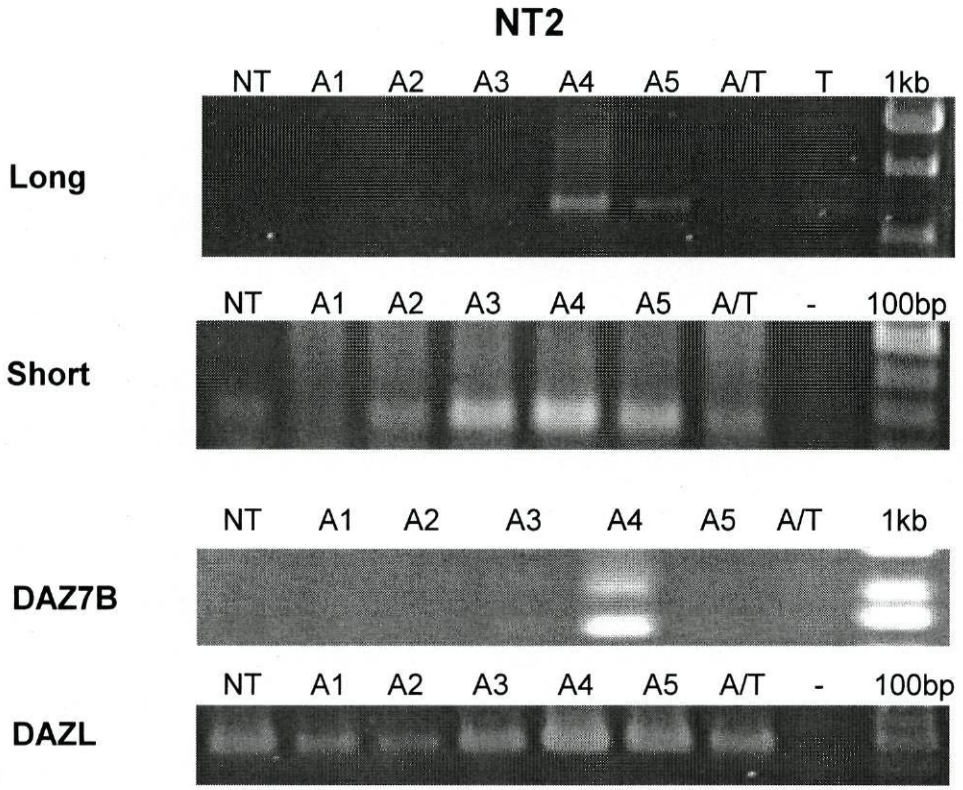


Figure 34 – DAZ transcripts on NT2 cell line after azacytidine treatments. NT – Testicular cancer cell line NT2; A1-A5 - Azacytidine treatments from 1 to 5 days; A/T – Azacytidine (day 2) + Trichostatin ; 1kb ladder or 100bp ladder used as a molecular weight marker.

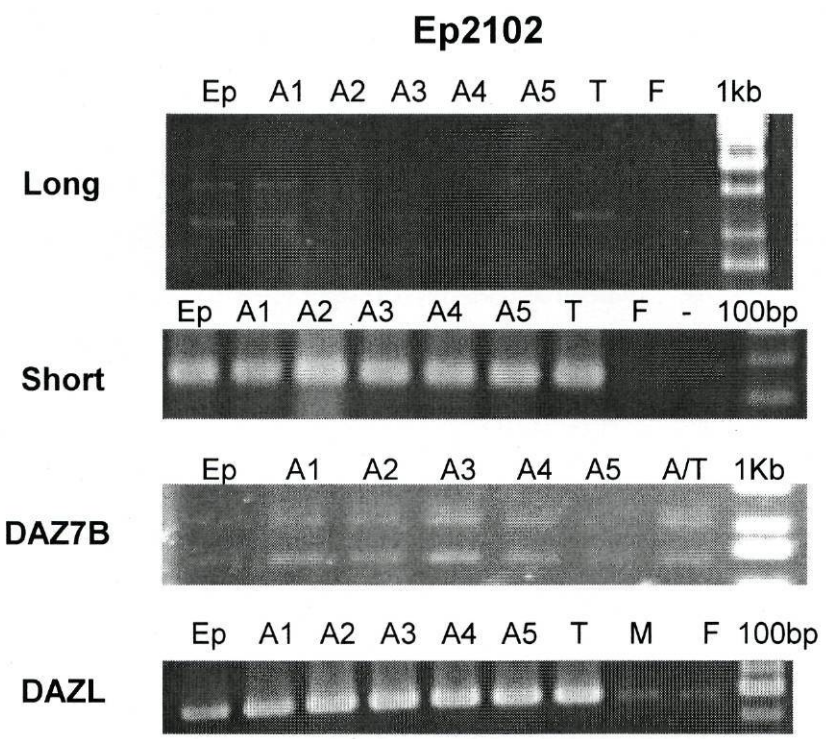


Figure 35 – DAZ transcripts on Ep2102 cell line after azacytidine treatments. Ep – Testicular cancer cell line Ep2102; A1-A5 - Azacytidine treatments from 1 to 5 days; A/T – Azacytidine (day 2) + Trichostatin ; 1kb ladder or 100bp ladder used as a molecular weight marker.

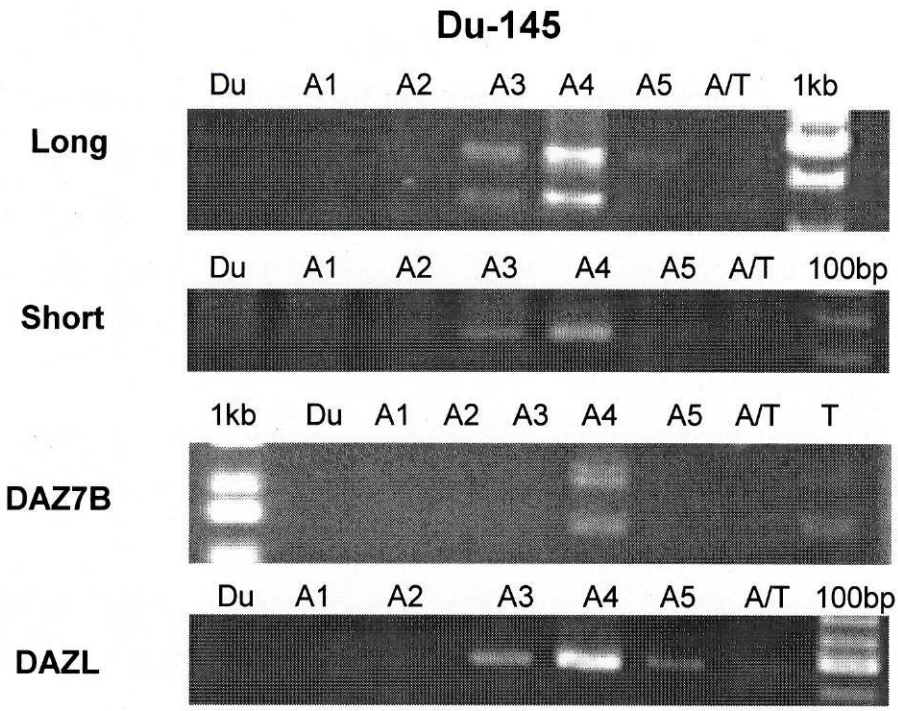


Figure 36 – DAZ transcripts on Du-145 cell line after azacytidine treatments. Du - Prostate cell line Du-145; A1-A5 - Azacytidine treatments from 1 to 5 days; A/T – Azacytidine (day 2) + Trichostatin ; 1kb ladder or 100bp ladder used as a molecular weight marker.

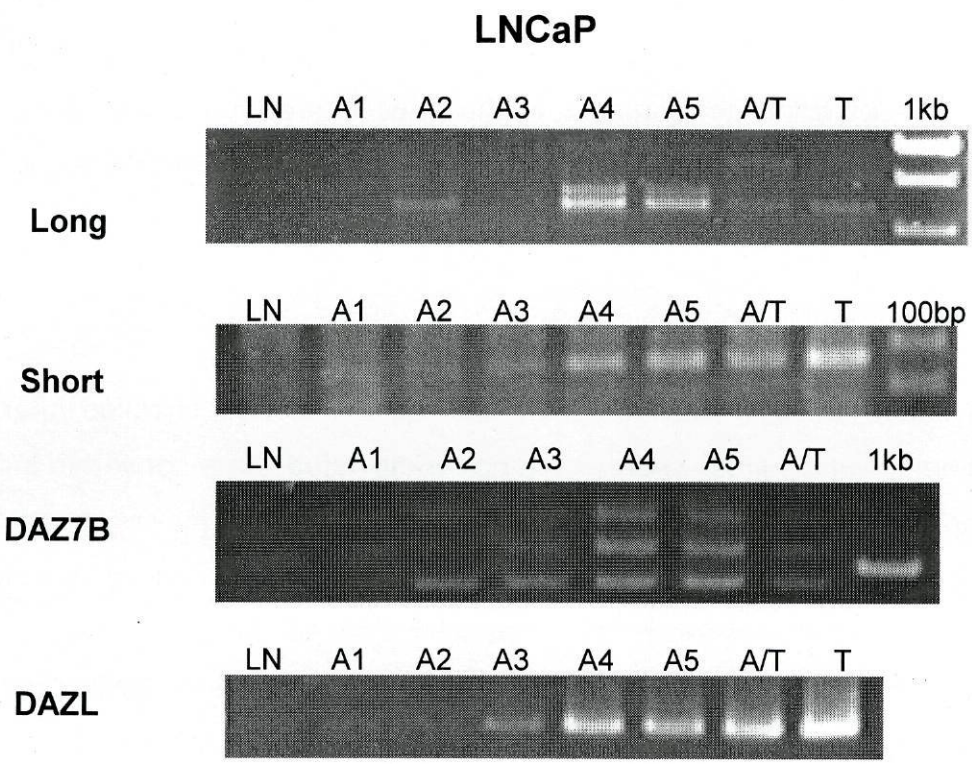


Figure 37 – DAZ transcripts on LNCap cell line after azacytidine treatments. LN - Prostate cell line LNCap; A1-A5 - Azacytidine treatments from 1 to 5 days; A/T – Azacytidine (day 2) + Trichostatin ; 1kb ladder or 100bp ladder used as a molecular weight marker.

The NT2 studies revealed that only after the third day of treatment it was possible to obtain long transcripts (only present on DAZ1 and DAZ4) as well as 7B transcripts (DAZ1, 3, 4). Short transcripts were detected after second day of treatment and this could be explained because all *DAZ* copies will produce this transcript. *DAZL* transcription activity seem to be methylation independent.

The Ep2102 cell line seem to be demethylated because the transcriptional activity of *DAZ* genes is independent of the treatment.

The *DAZ* transcriptional activity of Du-145 and LNCap cell lines presents a similar pattern to the one obtained with NT2 cell line, with the increase of transcription from the third for the fourth day being clearer.

The different cell lines displayed similar increase of transcriptional *DAZ* activity after on going demethylation. It was therefore possible to demonstrate that the transcriptional activity of *DAZ* genes is controlled by methylation and can be activated after a progressive loss of DNA methylation in each round of cell division (A1 to A5).

After the Trichostatin A (TCA) treatment, there was a decrease on the transcriptional activity in NT2, Du-145 and LNCaP. This can be indicative that histone hyperacetylation could modulate the gene expression, as referred by Yoshida (1995), although the TCA toxicity is also a factor that should be considered.

4.7. Bisulfite treatment

The methylation status of each CpG of CpG islands can be checked after a DNA treatment with bisulfite, whereas the non-methylated C are converted in T. Methylation specific PCR (MSP) uses the sequence differences between methylated and unmethylated alleles, which occur after sodium bisulfite treatment.

However, the bissulfite treatment performed in the four cell lines DNA did not produce conclusive results. This fact happens after several attempts to get a good sequence of the PCR product after bisulfite treatment. In the future, other experiments need to be developed in order to check if and which CpG doublet in the *DAZ* promoter CpG island are, indeed, methylated.

4.8. Transcriptional analyses of *DAZ*-L1 element

A complete LINE element was mapped in each *DAZ* copy between the 7X/7Y repeat in *DAZ* 1, 2 and 4 copies and in *DAZ* 3 between the 7X/7Z repeat (fig. 15). The sequence used to map the Line element on *DAZ* genes was the Human transposon L1.2 (GenBank ac. nr. M80343, Santos *et al.* 2000). L1.2 has 2 ORFs (Open Reading Frames) with the first one (ORF1) coding for a 40 kDa RNA binding protein and the second one (ORF2) coding for a protein containing an endonuclease and a reverse transcriptase domains. In *DAZ*-LINE, the sequence of ORF1 shows nucleotides changes that leads to two stop codons, in positions 330 and 390 of the ORF1 translated sequence.

Different promoter prediction programmes were used and a promoter region was found in *DAZ*-LINE element. In each *DAZ* copy gene the promoter and the ORFs have an inverted orientation in relation to the *DAZ* gene promoter.

The attempt to find specific *DAZ*-Line primers in order to check a putative transcriptional activity of this element within the *DAZ* gene was, however, not successful. This fact was due to the high homologies present on almost all the LINE sequences spread all over the genome.

Discussion

AZFc deletions are the most common known genetic cause of human male infertility and are found in different human populations with a rate of about 10^{-4} (Vogt 1996). In spite of this, it is not yet known whether the Y gene families mapped to this chromosomal region (fig. 4) and exclusively expressed in testicular tissue (Lahn & Page 1997, Vogt 1998, Kuroda-Kawaguchi *et al.* 2001) are indeed essential for human spermatogenesis. Gene specific deletions or point mutations have not been reported for any of the proposed AZFc candidate genes. Therefore it has been proposed that the observed AZFc testicular pathology might only probably be caused by deletion of the entire AZFc gene content of the human Y chromosome (Vogt 1998).

5.1. *AZFc* structure

The AZFc region is now covered with a contig of 44 BAC sequences and a molecular extension of about 4.94 Mb. A similar BAC contig of the AZFc region was recently published by Kuroda-Kawaguchi *et al.* (2001). Their data confirm the correct assembly of our AZFc-BAC contig shown on table VII. The same group extended this analysis by additional interactive alignment processes that revealed a unique repetitive and palindromic AZFc sequence structure along 93% of its length (Kuroda-Kawaguchi *et al.* 2001).

As the main conclusion for the first part of this work, strong experimental evidence was found showing that not only *AZFc* deletion but also partial *AZFc* deletions including the *DAZ*1/*DAZ*2 will cause the same pathology. It was found that in 5 of 63 oligozoospermic individuals (OZ), deletions of the *DAZ1/DAZ2* gene doublet are associated with a oligozoospermic phenotype similar to that observed after deletion of the entire AZFc region. These results are a major step towards the solution of the frequently discussed enigma on the function of the different *DAZ* gene copies in human spermatogenesis.

5.2. *DAZ*1/*DAZ*2 deletions associated with severe oligozoospermia

The number and structure of the *DAZ* genes on the human Y chromosome has been assumed to be variable, in a range between 3 and 7 gene copies (Yen *et al.* 1997, Gläser *et al.* 1998, Saxena *et al.* 2000). In the RPCI-11 BAC library, it was not found more than 4 *DAZ* gene copies nor the presence of truncated *DAZ*

pseudogenes. These 4 *DAZ* gene copies were mapped as *DAZ1*/*DAZ2* and *DAZ3*/*DAZ4* doublets in the middle of the 4.94 Mb long AZFc BAC contig with a distance of 1.47 Mb between them (Fig. 15). This genomic structure was first proposed in the AZFc duplication model by Kirsch *et al.* (1996). The presence of four *DAZ* genes with the same doublet structure was first found in the Y chromosomes of 8 men by Saxena *et al.* (2000), although with different repetitive exon 7 domains in *DAZ2*, *DAZ3* and *DAZ4*. The four copy gene structure found on the Y chromosome of the RPCI-11 donor was also found in 50 out of 107 fertile control individuals and in 32 out of 63 OZ samples (48%) of the samples analysed. Therefore, it has been proposed that the 4 *DAZ* gene pattern, found in the RPCI-11 *DAZ*-BACs, represents the most common *DAZ* gene structure in Caucasian male populations.

However, the *DAZ*-deletion analyses suggest that in some individuals (fertile men and oligozoospermic patients), the *DAZ* copy number is reduced to three by deletion of the *DAZ2* or *DAZ4* gene copy. The presence of only 3 *DAZ* gene copies on the Y chromosome of some fertile men was first proposed by Yen *et al.* (1997).

In DNA samples where the deletion of *DAZ2* or *DAZ4* could not be confirmed by appropriate DYS1-blot experiments, familial gene conversion events between the highly homologous *DAZ* gene copies (<99%) should have occurred. Gene conversion events, especially within the *DAZ2* and *DAZ4* gene structures, would easily explain the complex pattern of derivatives obtained for *DAZ*-del-hap 2, 4, and 9 in *DAZ*-SNV analyses. Similar conclusions were drawn recently for other human gene loci, based on unexpected linkage disequilibrium (LD) between closely spaced SNP markers (Ardlie *et al.* 2001). Obviously, gene conversion events prevent the identification of a specific *DAZ* copy deletion by only SNV/STS analyses and the additional analysis by DYS1-blot experiments is required.

A possible link of a specific *DAZ*-deletion haplotype specific to the occurrence of the patients' oligozoospermic phenotype was found only for the oligozoospermic samples with *DAZ*-del-hap 5 (5 of 63), pointing to a deletion of the *DAZ1* and *DAZ2* gene copies. These results could be confirmed by the DYS1-DNA blot analysis. *DAZ*-del-hap 5 was not found in any of the 107 fertile men samples. It

is, then, assumed that the *DAZ1*/*DAZ2* deletions are most likely *de novo* mutation events, restricted to the *DAZ* locus of each patient.

The analyses of *DAZ* gene copy specific restriction fragments in the RPCI-11 *DAZ*-BAC contigs were only successful for the repetitive DYS1-*DAZ* gene regions. The exact molecular extensions of the *DAZ*1/*DAZ*2 gene deletions could not be estimated. However, absence of the *DAZ*1 specific DYS1-*Eco*RV fragments in all *DAZ*-del-hap 5 patients includes a deletion of at least two of the *DAZ*1 RNA binding domains (RRM) in exon 2-5 (Fig. 38). A distortion of the putative *DAZ*1 germ line specific RNA binding function is then expected.

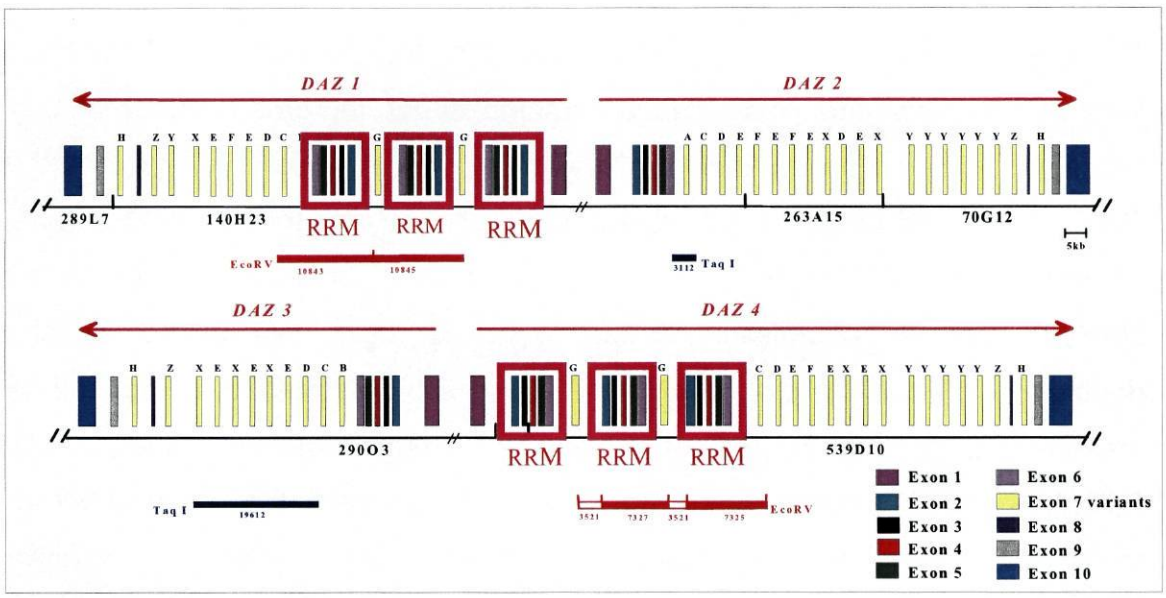


Figure 38 - *DAZ* genes with the 3 RRM domains on *DAZ*1 and *DAZ*4 copies. Below each copy, the position of DYS1-*Eco*RV and *Taq*I specific fragments.

5.3. *DAZ3*/*DAZ4* deletions in fertile men

The highly repeated nature of the sequences within AZFc required the establishment of assays for specific variants within the *DAZ* regions that allowed to distinguish among similar sequences. These markers were extended to other regions in the AZFc region, providing an assay for 8 loci spread over 3Mb in this region.

It was shown that Y chromosomes carrying a large AZFc deletion that removes the *DAZ*3 and *DAZ*4 genes, are common in the European population. It was shown here that haplogroup N chromosomes carry this deletion. This haplogroup is common throughout Northern Europe and Asia, making up to

12% of Y-chromosomes in one world-wide survey (Zerjal *et al.* 1997). It is an ancient and successful lineage where the lack of *DAZ*3 and *DAZ*4 has no detrimental effect. This deletion has existed for a long time in normal males, so it was possible to conclude that *DAZ*3 and *DAZ*4 gene copies are not essential for spermatogenesis, at least in these men.

5.4. Putative molecular mechanisms for the *DAZ* genes deletion

The comparison between the *DAZ* gene structures of the RPCI-11 donor and that of the CTA/CTB donors described by Saxena *et al.* (2000) reflects a dynamic structure of the *DAZ*2, *DAZ*3 and *DAZ*4 gene copies but not of the *DAZ*1 gene copy. The *DAZ* gene doublets *DAZ1*/*DAZ2* and *DAZ*3/*DAZ*4 are located in two palindromic sequence domains, separated by a distance of 1.47 Mb (figure 23). The observed dynamic rearrangements, which distinguished the *DAZ*2, *DAZ*3 and *DAZ*4 gene copies of the CTA/CTB and the RPCI-11 donors, as well as gene conversion events (Ardlie *et al.* 2001) might be caused by the structural specific organisation in the genomic region of both *DAZ* gene doublets and could be extended to the entire AZFc sequence, composed by massive palindromes in 93% of its length (Kuroda-Kawaguchi *et al.* 2001). Palindromes and inverted repeat structures in the genome are generally known to induce local instabilities which depend on their lengths and repetitive sequence contents. Local rearrangements increasing their asymmetry could stabilise them (Zhou *et al.* 2001). In the *DAZ*1/*DAZ*2 and *DAZ*3/*DAZ*4 palindromes, this could explain the amplification of RRM domains in only *DAZ*1 and *DAZ*4 and the variable structure of the repetitive exon 7 in *DAZ*2, *DAZ*3 and *DAZ*4.

An important prerequisite for a recombination event is the existence of repetitive sequence blocks, presenting the same polarity. On figure 39 are presented the potential intrachromosomal recombination in AZFc locus between amplicons presenting >99.9% homology (arrows with the same colour) and the same polarity (same arrow direction).

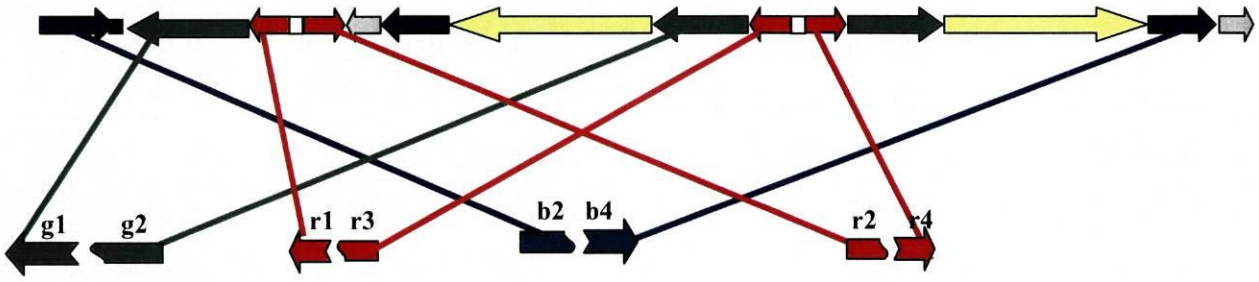


Figure 39 – Potential deletions in the AZFc locus due to intrachromosomal recombinations of four different amplicons present in AZFc with the same polarity (Vogt & Fernandes, 2003).

Table XXI shows the number of functional AZFc gene copies affected by each potential recombination event leading to the respective AZFc internal deletion.

	b2/b4	g1/g2	r1/r3	r2/r4
BPY2	0	2	2	2
DAZ	0	2	2	2
CDY1	0	1	1	1
CSPG4LY	0	1	1	1
GOLGA2LY	0	1	1	1
TTY3	0	1	1	1
TTY4	0	1	1	1

Table XXI – Number of functional AZFc gene copies present after recombinations and internal AZFc deletions. (According to Yen 2001).

After analysing the palindromic structure of *AZFc* region and each recombinant event, the gene content of this region can be reduced to zero, when a complete AZFc deletion happen (b2/b4), but it will be only partially reduced when other recombination mechanism occur (g1/g2, r1/r3 or r2/r4). These findings, together with the fact that the Y chromosome sequence is now totally known, enable, in a near future, to characterize the breakpoints and to identify the exact genes deleted in each oligozoospermic patient presenting partial AZFc microdeletions.

5.4.1. Mechanism for *DAZ*1/*DAZ*2 genes deletion

Because deletions of the *DAZ*1 gene alone were not found but only in combination with the *DAZ*2 gene, it was possible to assume that molecular mechanisms inducing the observed *DAZ*1/*DAZ*2 gene deletion events are most likely based on the recombination of homologous repetitive sequence blocks flanking the *DAZ*1/*DAZ*2 gene doublet (Chen *et al.* 1997).

Intrachromosomal recombination events of two repetitive 10 kb sequence blocks in proximal Yq11 were also found as causative agents for the occurrence of most AZFa deletions (Kamp *et al.* 2000; Sun *et al.* 2000). Long homologous repetitive sequence blocks were also found around the *DAZ* genes.

Analysis of the AZFc palindromic structure shows the existence of two long repetitive blocks in the same polarity (g1 and g2 amplicons) and with a high sequence homology, extending to the 3´ends of the *DAZ*1 and *DAZ*3 gene respectively. Intrachromosomal recombination events between these two green amplicons would delete not only the *DAZ*1 gene but also the entire *DAZ*2 gene.

A deletion of the complete *DAZ*1/*DAZ*2 gene doublet could be caused by intrachromosomal recombination events between the b1 and b3 amplicons or, as a second possibility, between the AZFc amplicons g1 and g2.

The exclusion of a b1/b3 recombination event as causative agent for the complete deletion of the DAZ1/DAZ2 gene doublet in the samples with DAZ-del-hap 5 was assumed by the presence of two STSs, namely DYS227/sY139 (GenBank ac. no. G12002), mapped to AZFc-u2 and sY1192 (GenBank accession no. G67166), mapped to AZFc-u3.

A *DAZ*1/*DAZ*2 deletion caused by g1/g2 recombination event is then the most probable mechanism for the occurrence of the oligozoospermic phenotype in samples with the *DAZ*-del-hap 5. This putative molecular mechanism is shown in figure 40.

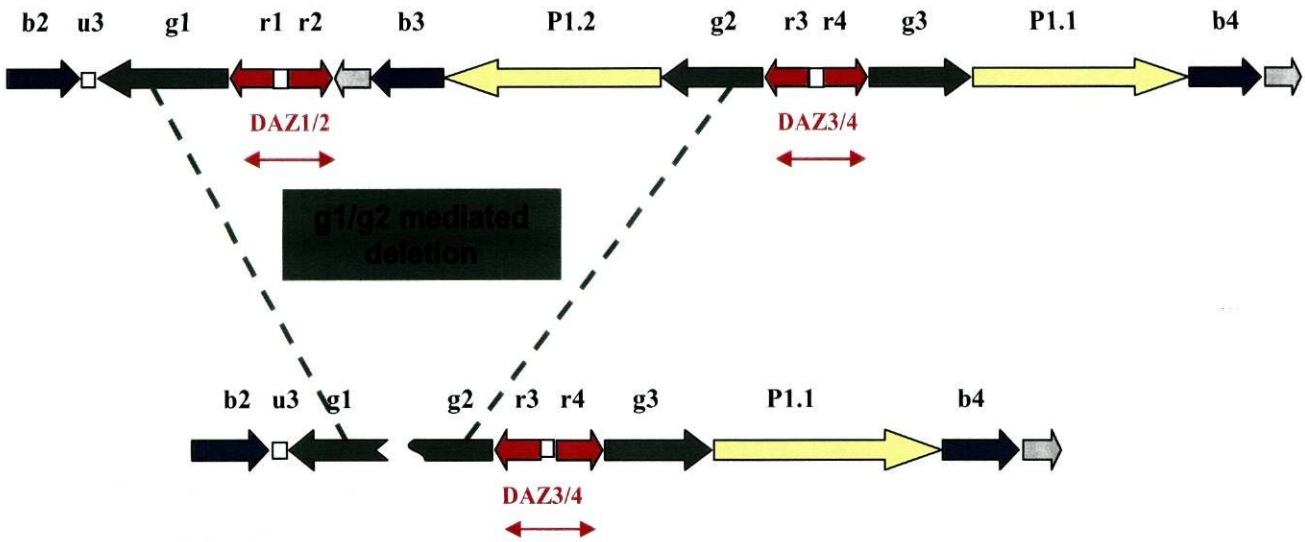


Figure 40 – Putative molecular recombinant mechanism responsible for deletion of *DAZ*1/*DAZ*2 doublet.

5.4.2. *DAZ*3/DAZ4 genes deletion

Complete *AZFc* deletions were described as a intrachromosomal recombination event between the b2 and b4 amplicons (Kuroda-Kawaguchi *et al.* 2001) and deletions of the *DAZ*1/*DAZ*2 gene doublet occurring through recombination events between the g1 and g2 amplicons (Fernandes *et al.* 2002). From the *AZFc*-SNVs analyses combined with the Y-haplogroup studies performed with fertile men samples, it was possible to associate a deletion of the complete *DAZ*3/*DAZ*4 gene doublet in samples belonging to Y-haplogroup N with the absence of the sY1192 and 50f2/C markers, located in *AZFc*-u3.

There are no obvious candidates for directly repeated amplicons flanking the two deleted regions, so recombination between long homologous flanking sequences cannot provide an explanation for the DAZ3/DAZ4 deletions in the N lineage Y chromosomes. This deletion includes two AZFc regions that are non-contiguous in the GenBank sequence (Fig. 24) and so cannot readily be explained in a simple way. Two more complex mechanisms could be proposed: (1) two independent deletions could have occurred, one removing u3 and other the g2-DAZ3-DAZ4-g3-P1.1 region; alternatively (2) the deletion substrate could differ from the GenBank AZFc sequence and be rearranged in the Y chromosomes of the pre-N lineage, carrying u3 adjacent to the g2-DAZ3-DAZ4-g3-P1.1 region. In that case, a single deletion event would remove all of the

sequence missing in the N lineage by a similar intrachromosomal recombination event to that seen for DAZ1/DAZ2 and complete AZFc deletions (fig. 39).

While it is possible that short regions of sequence identity, not apparent in the gross AZFc amplicon structure, might be involved in the DAZ3/DAZ4 deletions, it has been explored the alternative possibility that evolutionarily earlier rearrangements in the AZFc amplicons could have produced a structure that differs from the GenBank sequence and provides a substrate for a single deletion event.

Such rearrangements in the GenBank AZFc structure could place u3 adjacent to the g2-DAZ3-DAZ4-g3-P1.1 region: Inversion between the b2 and b3 amplicons followed by inversion between the P1.1 and P1.2 or b2 and b4 amplicons or vice versa leading to the same end product where then, a single g1/g2 recombination would delete all the sequences absent from lineage N chromosomes in a single step (Fig. 41). In such AZFc amplicon structure, typical for N lineage Y chromosomes, the putative mechanism responsible for the complete deletion of AZFc region will involve an intrachromosomal mechanism event between the b3 and b4 amplicons.

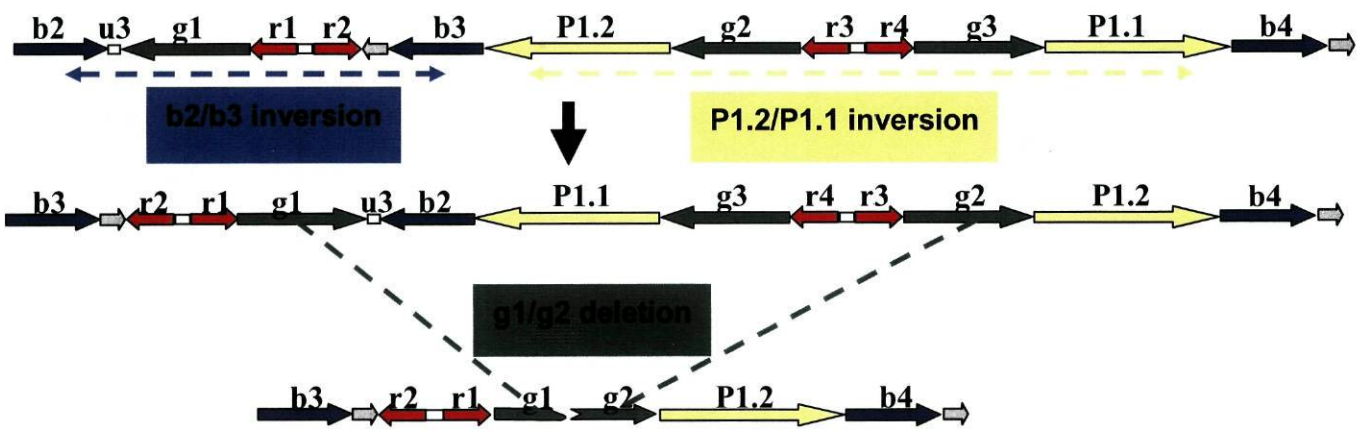


Figure 41 – Putative molecular recombinant mechanism responsible for deletion of *DAZ3*/*DAZ4* doublet.

5.5. *DAZ* promoter

DAZ promoter studies were not conclusive and other tests need to be performed in order to check if the predicted sequences obtained by *in silico* analysis have indeed a promoter activity. This functional analysis could be done by subcloning the intronic region between exon 1 and exon 2 from each *DAZ* gene copy in an expression vector suitable for the analyses of functional promoter elements, like pGL3-Luciferase® system from Promega.

5.6. *DAZ* transcriptional regulation

The methylation studies performed in this work led to the assumption that *DAZ* gene present a transcriptional activity regulated by methylation. The results obtained by using testicular and prostate cell lines indicated that the first gene to be transcribed is *DAZ*2, at least in 2 cell lines, and that only 3 *DAZ* gene copies are transcribed in all cell lines, excepting LNCaP. These findings and the fact that different *AZFc* composition associates with different Y-haplogroups, suggest that LNCaP should belong to a different Y-haplogroup, where all 4 *DAZ* copies are present and transcribed. However, further research projects need to be developed in order to check this hypothesis. Real-time PCR technology is now available and could be used to quantify, separately, the transcription of each *DAZ* gene, clarifying whether all *DAZ* genes became activated simultaneously or at different times.

5.7. Conclusion

Molecular biology has revealed that microdeletions of the Y chromosome represent an important cause of male infertility and the most frequent genetic aetiology of severe impairment of spermatogenesis. Such findings are fundamental for a careful diagnosis of male infertility and its treatment. The Y chromosome deletion screening is now a reality in the majority of infertility centres. The detection of a deletion in an infertile man provides a proper diagnosis of the disease, avoiding unnecessary and often expensive treatments to improve fertility (*e.g.* hormonal treatments), and has important ethical consequences if the patient is a candidate for assisted reproduction techniques.

The identification of the actual role played by the *AZFc* candidate genes in spermatogenesis will provide significant advances to a better understanding of the biology of spermatogenesis and the analysis of Y-chromosomal genes with a potential role in male germ cell development.

With the establishment of a set of *DAZ*-SNVs/STSs is now possible to distinguish each *DAZ* gene copy and to detect, for the first time, partial *DAZ* gene deletions in a systematic way. Because gene conversion events exist, an additional analysis by DYS1 blot experiments was developed to confirm the identification of a specific *DAZ* copy deletion achieved by SNV/STS analyses.

The establishment of a *AZFc*-SNVs set extended this analysis to entire *AZFc* region, and revealed to be essential for a better understanding of recombinant mechanisms responsible for *AZFc* partial deletions

There are also some clinical implications from this work. After the detailed description of the genes responsible for some cases of infertility, the basis for molecular deletion diagnosis of each *AZFc* gene copy is now provided. This increases the possibilities of genetic counselling, giving clues about the father-to-son transmission and increasing the preimplantation genetic diagnosis (PIGD) possibilities.

An understanding of the structure and function of the *AZFc* genes, in future research projects, and the knowledge of their interactions will facilitate the development of potential therapies. The development of drugs able to restore the normal proceed of spermatogenesis will be only possible after understanding the function and the pathways where these genes are involved.

The Y chromosome is unique under many aspects. It is always in the haploid state, is full of repeated sequences but is responsible for important biological roles such as male sex determination and male fertility.

For all these reasons the Y chromosome is an important component of the human genome, at least for about 50% of all individuals in all populations - the so called men.

References

Agulnik AI, Zharkikh A, Boettger-Tong H, Bourgeron T, McElreavey K, Bishop CE (1998): Evolution of the *DAZ* gene family suggests that Y-linked *DAZ* plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. Hum. Mol. Genet., **7(9)**: 1371-1377.

Andrews P (1984): NT2 and 2102Ep cell lines. Sheffield, UK.

Ardlie K, Liu-Cordero SN, Eberle MA *et al.* (2001): Lower-than-expected linkage disequilibrium between tightly linked markers in human suggests a role for gene conversion. Am. J. Hum. Genet., **69**: 582-589.

Balhorn R (1982): A model for the structure of chromatin in mammalian sperm. J. Cell Biol., **93** (2): 298-305.

Bardoni B, Zuffardi O, Guioli S, Ballabio A, Simi P, Cavalli P, Grimoldi G, Fraccaro M, Camerino G (1991): A deletion map of the human Yq11 region: Implications for the evolution of the Y chromosome and tenative mapping of a locus involved in spermatogenesis. Genomics, **11**: 443-451.

Bird AP (1992): The essentials of DNA methylation. Cell, **70**: 5-8.

Bishop C, Guellaen G, Geldwerth D, Fellous M, Weissenbach J (1984): Extensive sequence homologies between Y and other human chromosomes. J. Mol. Biol., **173**: 403-417.

Birnboim HG, Doly J (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acid. Res., **7**: 1513-1523.

Broday L (1999): 5-azacytidine induces transgene silencing by DNA methylation in Chinese hamster cells. Mol. Cell Biol., **19**: 3198-3204.

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

Burgoyne PS (1996): Friut(less) flies provide a clue. Nature, **381** (6585): 740-741.

Burgoyne PS (1998): The mammalian Y chromosome: a new perspective. BioEssays, **20**: 363-366.

Charlesworth B (1991): The evolution of sex-chromosomes. Science, **251**: 1030-1033.

Chen B, Kung HF, Bates RR (1976): Effects of methylation of the beta-galactosidase genome upon *in vitro* synthesis of beta-galactosidase. Chem. Biol. Interact., **14**: 101-111.

Chen K-S, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR (1997): Homologous recombination of a flanking repeat gene cluster is a mechanism for common contiguous gene deletion syndrome. Nature Genet., **17**: 154-163.

Christensen ME, Rattner JB, Dixon GH (1984): Hyperacetylation of histone H4 promotes chromatin decondensation prior to histone replacement by protamines during spermatogenesis in rainbow trout. Nucl. Acids Res., **12** (**11**): 4575-4592.

Clermont Y (1966): Renewal of spermatogonia in man. Am. J. Anat., **118**: 509-524.

Cooke HJ, Lee M, Kerr S, Ruggiu M (1996): A murine homologue of the human *DAZ* gene is autosomal and expressed only in male and female gonads. Hum. Mol. Genet., **5**: 513-516.

Dym M (1994): Spermatogonial stem cells of the testis. Proc. Natl. Acad. Sci., **91**: 11287-11289.

Eberhart CG, Maines JZ, Wasserman SA (1996): Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. Nature **381** (6585): 783-785.

Edwards RG & Bishop CE (1997): On the origin and frequency of Y chromosome deletions responsible for severe male infertility. Mol. Hum. Reprod. **3**: 549-554.

Fanning TG, Singer MF (1987): LINE-1: a mammalian transposable element. Biochim. Biophys. Acta, **910**: 203-212.

Ferlin A, Moro E, Rossi A, Foresta C (2002): A novel approach for the analysis of DAZ gene copy number in severely idiopathic infertile men. J. Endocrinol. Invest., **25**: RC1-RC3.

Feng Q, Moran JV, Kazazian HH, Boeke JD (1996): Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell, **87**: 905-916.

Fernandes S, Huellen K, Gonçalves J, Dukal H, Zeisler J, Rajpert De Meyts E, Skakkebaek NE, Habermann B, Krause W, Sousa M, Barros A, Vogt PH (2002): High frequency of *DAZ1/DAZ2* gene deletions in patients with severe oligozoospermia. Mol. Hum. Reprod., **8** (**3**): 286-298.

Fickett JW, Hatrigeorgion AG (1997): Eukaryotic promoter recognition. Genome Res., **7(9)**: 861-878.

Frengen E, Zhao B, de Jong PJ (1999): A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. Genomics, **58**: 250-253.

Gläser B, Yen P, Schempp W (1998): Fibre-fluorescence *in situ* hybridisation unravels apparently seven *DAZ* genes or pseudogenes clustered within a Y-chromosome region frequently deleted in azoospermic males. Chrom. Res., **6**: 481-486.

Gromoll J, Weinbauer GF, Skaletsky H, Schlatt S, Rocchietti-March M, Page DC, Nieschlag E (1999): The old world monkey *DAZ* (deleted in azoospermia) gene yields insights into the evolution of the *DAZ* gene cluster on the human Y chromosome. Hum. Mol. Genet., vol. **8 (11)**: 2017-2024.

Grunau C, Hindermann W, Rosenthal A (2000): Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes. Hum. Mol. Genet., **9 (18)**: 2651-2663.

Habermann B, Mi HF, Edelmann A, Bohring C, Backert IT, Kiesewetter F, Aumüller G, Vogt PH (1998): *DAZ* (Deleted in Azoospermia) genes encode proteins located in human late spermatids and in sperm tails. Hum. Reprod., **13 (2)**: 363-369.

Hecht NB (1995): The making of a spermatozoon: A molecular perspective. Dev. Genet., **16**: 95-103.

Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH (1997): Multiplex PCR: Critical parameters and step-by-step protocol. BioTechniques, **23**: 504-511.

Jobling MA, Williams GA, Schiebel K, Pandya A, McElreavey K, Salas L, Rappold GA, Affara NA, Tyler-Smith C (1998): A selective difference between human Y-chromosomal DNA haplotypes. Curr. Biol., **8**: 1391-1394.

Jost JP (1993): Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. Proc. Natl. Acad. Sci. USA, **90**: 4684-4688.

Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH (2000): Two homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. Hum. Mol. Genet., **9**: 2563-2572.

Kennedy BP, Davies PL (1980): Acid-soluble nuclear proteins of the testis during spermatogenesis in the winter flounder. Loss of the high mobility group proteins.J. Biol. Chem., **255** (**6**): 2533-2539.

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

Keshet I, Lieman-Hurwitz J, Cedar H (1986): DNA methylation affects the formation of active chromatin. Cell, **44**: 535-543.

Knudser S (1999): Promoter2.0: for recognition of Pol II Promoter sequences. Bioinformatics, **15**: 356-361.

Kondrakhin YV, Kel AE, Kolchanov NA, Romashchenko AG, Milanesi L (1995): Eukaryotic promoter recognition by binding sites for transcription factors. CABIOS. **11**: 477-488.

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC (2001): The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nature Genetics, **29**: 279-286.

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

Lahn BT & Page DC (1999): Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. Nat. Genet., **21**: 429-433.

Lahn BT, Tang ZL, Zhou J, Barndt RJ, Parvinen M, Allis CD, Page DC (2002): Previously uncharacterised histone acetyltransferases implicated in mammalian spermatogenesis. PNAS, **99** (**13**): 8707-8712.

Lau Y-F & Zhang J (2000): Expression analysis of thirty one Y chromosome genes in human prostate cancer. Mol. Carcinogenesis, **27**: 308-321.

Ma K, Inglis J, Sharkey A, Bickmore W, Hill R, Prosser E, Speed R, Thomson E, Jobling M, Taylor K, Wolfe J, Cooke H, Hargreave T, Chandley A (1993): A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. Cell, **75**: 1287-1295.

Ma K, Mallidis C, Bhasin S (2000): The role of Y chromosome deletions in male infertility. Eur. J. Endocrinol., **142**: 418-430.

Maiwald R, Seebacher T, Edelmann A, Hirschmann P, Köhler MR, Kirsch S, Vogt PH (1996): A human Y-chromosomal gene mapping to distal Yq11 is expressed in spermatogenesis. Report of the 2nd International Workshop on Y Chromosome mapping 1995. Cytogenet. Cell Genet., **73**: 33-76.

Manniatis T, Fritsch EF, Sambrook J (2001): Molecular Cloning: A laboratory manual. 3rd ed., Cold Spring Harbor Laboratory Press, New York, USA.

Mathias SL, Scott AF, Kazazian HH, Boeke JD, Gabriel A (1991): Reverse transcriptase encoded by a human transposable element. Science, **254**: 1808-1810.

Mazeyrat S, Saut N, Sargent CA, Grimmond S, Longepied G, Ehrmann IE, Ellis PS, Greenfield A, Affara NA, Mitchell MJ (1998): The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region. Hum. Mol. Genet., **7**: 1713-1724.

Meistrich ML, Trostle-Weige PK, Lin R, Bhatnagar YM, Allis CD (1992): Highly acetylated H4 is associated with histone displacement in rat spermatids. Mol. Reprod. Dev., **31 (3)**: 170-181.

Menke DB, Mutter GL, Page DC (1997): Expression of *DAZ*, an azoospermia factor candidate, in human spermatogonia. Am. J. Hum. Genet., **60**: 237-241.

Miller SA, Dykes DD, Polesky HF (1988): A simple salting out procedure for extracting DNA from human nucleated cells. Nucl. Acid. Res., **16 (3)**: 12-15.

Morton NE (1991): Parameters of the human genome. Proc. Nat. Acad. Sci., **88**: 7470-7476.

Ng HH & Bird A (1999): DNA methylation and chromatin modification. Curr. Op. In Genet. & Development, **9**: 158-163

Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986): A DNA probe detecting multiple haplotypes of the human Y chromosome. Am. J. Hum. Genet., **38**: 407-418.

Paracchini S, Arredi B, Chalk R, Tyler-Smith C (2002): Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. Nucl. Acids Res., **30**: e27.

Parthun MR, Widom J, Gottschling DE (1996): The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. Cell, **87**: 85-94.

Périer RC, Junier T, Bucher P (1998): The Eukaryotic Promoter Database EPD. Nucleic Acids Res., **26**: 353-357.

Pierce JC, Sauer B, Sternberg N (1992): A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. Proc. Natl. Acad. Sci., USA, **89**: 2056-2060.

Prestridge DS (1995): Predicting Pol II promoter sequences using transcription factor binding sites. J. Mol. Biol., **249**: 923-932.

Quandt K, Frech K, Karas H, Wingender E, Werner T (1995): MatInd and MatInspector – new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res., **23**: 4878-4884.

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

Reijo R, Seligman J, Dinulos MB, Jaffe T, Brown LG, Disteche CM, Page DC (1996): Mouse autosomal homolog of *DAZ*, a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty. Genomics, **35**: 346-352.

Rideout WMI, Coetzee GA, Olumi AF, Jones PA (1990): 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and *p*53 genes. Science, **249**: 1288-1290.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E *et al.* (2000): Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet., **67**: 1526-1543.

Ruggiu M, Speed R, Taggart M, McKay SJ, Kilanowski F, Saunders P, Dorin J, Cooke HJ (1997): The mouse *DAZla* gene encodes a cytoplasmic protein essential for gametogenesis. Nature, **389**: 73-77.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985): Enzymatic amplification of ß-Globin genomic sequences and restriction site analysis for diagnosis of Sickle Cell Anemia. Science, **230**: 1350-1354.

Santos FR, Pandya A, Kayser M, Mitchell RJ, Liu A, Singh S, Destro-Bisol G, Novelletto A, Qamar R, Mehdi SQ, Adhikar R, de Knijff P, Thyler-Smith C (2000): A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome. Hum. Mol. Genet., **9** (**3**): 421-430.

Sargent CA, Boucher CA, Kirsch S, Brown G, Weiss B, Trundley A, Burgoyne P, Saut N, Durand C, Levy N, Terriou P, Hargreave T, Cooke H, Mitchell M, Rappold GA, Affara NA (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.

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

Saxena R, de Vries JWA, Repping S, Alagappan RK, Skaletsky H, Brown LG, Ma P, Chen E, Hoovers JMN, Page DC (2000): Four *DAZ* Genes in Two Clusters Found in the AZFc Region of the Human Y Chromosome. Genomics, **67**: 256-267.

Schaffer HE & Sederoff RR (1981): Improved estimation of DNA fragment lengths from agarose gels. Anal. Biochem., **115**(**1**): 113-122.

Scott AF (1987): Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics, **1**: 113-125.

Shan Z, Hirschmann P, Seebacher T, Edelmann A, Jauch A, Morell J, Urbitsch P, Vogt PH (1996): A *SPGY* copy homologous to the mouse gene *Dazla* and the *Drosophila* gene *boule* is autosomal and expressed only in the human male gonad. Hum. Mol. Genet., **5** (**12**): 2005-2011.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992): Cloning and stable maintenance of 300 kilobase pair fragments of human DNA in *Escherichia coli* using na F-factor based vector. Proc. Natl. Acad. Sci. USA, **89**: 8794-8797.

Simoni M, Bakker E, Eurlings MCM, Matthijs G, Moro E, Müller CR, Vogt PH (1999): Laboratory guidelines for molecular diagnosis of Y-chromosomal microdeletions. Int. J. Androl., **22**: 292-299.

Skaletsky H, Kuroda-Kawaguchi T, Minx P, Cordum HS, Brown L, Repping S, *et al.* (2003): The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature, **423**: 825-837.

Slee R, Grimes B, Speed RM, *et al.* (1999): A human *DAZ* transgene confers partial rescue of the mouse *DAZl* null phenotype. Proc. Natl. Acad. Sci., **96**: 8040-8045.

Sousa M, Fernandes S, Barros A (2000): Os genes e o homem infértil. Andrologia clínica, 195-222.

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

Stuppia L, Gatta V, Fogh I, Gaspari AR, Grande R, Morizio E, Fantasia D, Pizzuti A, Calabrese G, Palka G (2000): Characterisation of novel genes in AZF regions. J. Endocrinol. Invest. **23**: 659-663.

Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC (1999): An azoospermic man with a *de novo* point mutation in the Y chromosomal gene USP9Y. Nat. Genet., **23**: 429-432.

Sun C, Skaletsky H, Roz S, Gromoll J, Nieschlag E, Oates R, Page DC (2000): Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. Hum. Mol. Genet., **9**: 2291-2296.

Swergold GD (1990): Identification, characterisation and cell specificity of a Human LINE-1 promoter. Mol. Cell. Biology, **10**(**12**): 6718-6729.

Tazi J & Bird A (1990): Alternative chromatin structure at CpG islands. Cell, **60**: 909-920.

Tiepolo L & Zuffardi O (1976): Localisation of factors controlling spermatogenesis in the non-fluorescent portion of the Y chromosome long arm. Hum. Genet., **34**: 119-124.

Underhill P *et al.* (1997): Detection of numerous Y chromosome biallelic polymorphisms by denaturing high performance liquid chromatography (DHPLC). Genome Res., **7**: 996-1005.

Underhill P, Shen P, Lin A, Jin L, Passarino G, Yang W, Kauffman E, Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd J, Mehdi S, Seielstad M, Wells R, Piazza A, Davis R, Feldman M, Cavalli-Sforza L, Oefner P (2000): Y chromosome sequence variation and the history of human populations. Nat. Genet., **26**: 358-361.

Unnerus V, Fellman J, De La Chapelle A (1967): The length of the human Y chromosome. Cytogenetics, **6**: 213-227.

Vogt PH, Chandley AC, Hargreave TB, Keil R, Ma K, Sharkey A (1992): Microdeletions in interval 6 of the Y chromosome of males with idiopathic sterility point to disruption of AZF, a human spermatogenesis gene. Hum. Genet., **89**: 491-496.

Vogt PH, Keil R, Kirsch S (1993): The AZF function of the human Y chromosome during spermatogenesis. In: Chromosomes Today, **11** (Summer, AT & Chandley AC, eds.). Chapman & Hall, London, 227-239.

Vogt PH (1996): Human Y chromosome function in male germ cell development. Adv. Dev. Biol., **4**: 191-257.

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Gröne H-J, Jung A, Engel W, Haidl G (1996): Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum. Mol. Genet., **5** (**7**): 933-943.

Vogt PH, Affara N, Davey P, Hammer M, Jobling MA, Lau Y-FC, Mitchell M, Schempp W, Tyler-Smith C, William G, Yen P, Rappold G (1997): Report of the third international workshop on Y-chromosome mapping 1997. Cytogenet. Cell Genet., **79**: 1-20.

Vogt PH (1998): Human chromosome deletion in Yq11, AZF candidate genes and male infertility: history and update. Mol. Hum. Reprod., **4**: 739-744.

Vogt PH, Fernandes S (2003): Polymorphic *DAZ* gene family in polymorphic structure of AZFc locus: Artwork or functional for human spermatogenesis? APMIS, **111**: 115-127.

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

Webber MM, Bello D, Quader S (1996): Immortalised and tumorigenic adult human prostatic epithelial cell lines: characteristics and applications. Prostate, **29**: 386-394.

Wolffe AP, Pruss D (1996): Targeting chromatin disruption: Transcriptional regulators that acetylate histones. Cell, **84**: 817-819.

Wouters-Tyrou D, Martinage A, Chevaillier P, Sautiere P (1998): Nuclear basic proteins in spermiogenesis. Biochimie, **80**: 117-128.

Wright S & Finnegan D (2001): Genome evolution: Sex and the transposable element. Curr. Biology **11**: R296-R299.

Yen PH, Chai NN, Salido EC (1997): The human *DAZ* genes, a putative male infertility factor on Y chromosome, are highly polymorphic in the *DAZ* repeat regions. Mamm. Genome, **8**: 756-759.

Yen PH, (1998): A long range restriction map of deletion interval 6 of the human Y chromosome: a region frequently deleted in azoospermic males. Genomics, **54**: 5-12.

Yen PH, (2001): The fragility of fertility. Nat. Genet., **29**: 243-244.

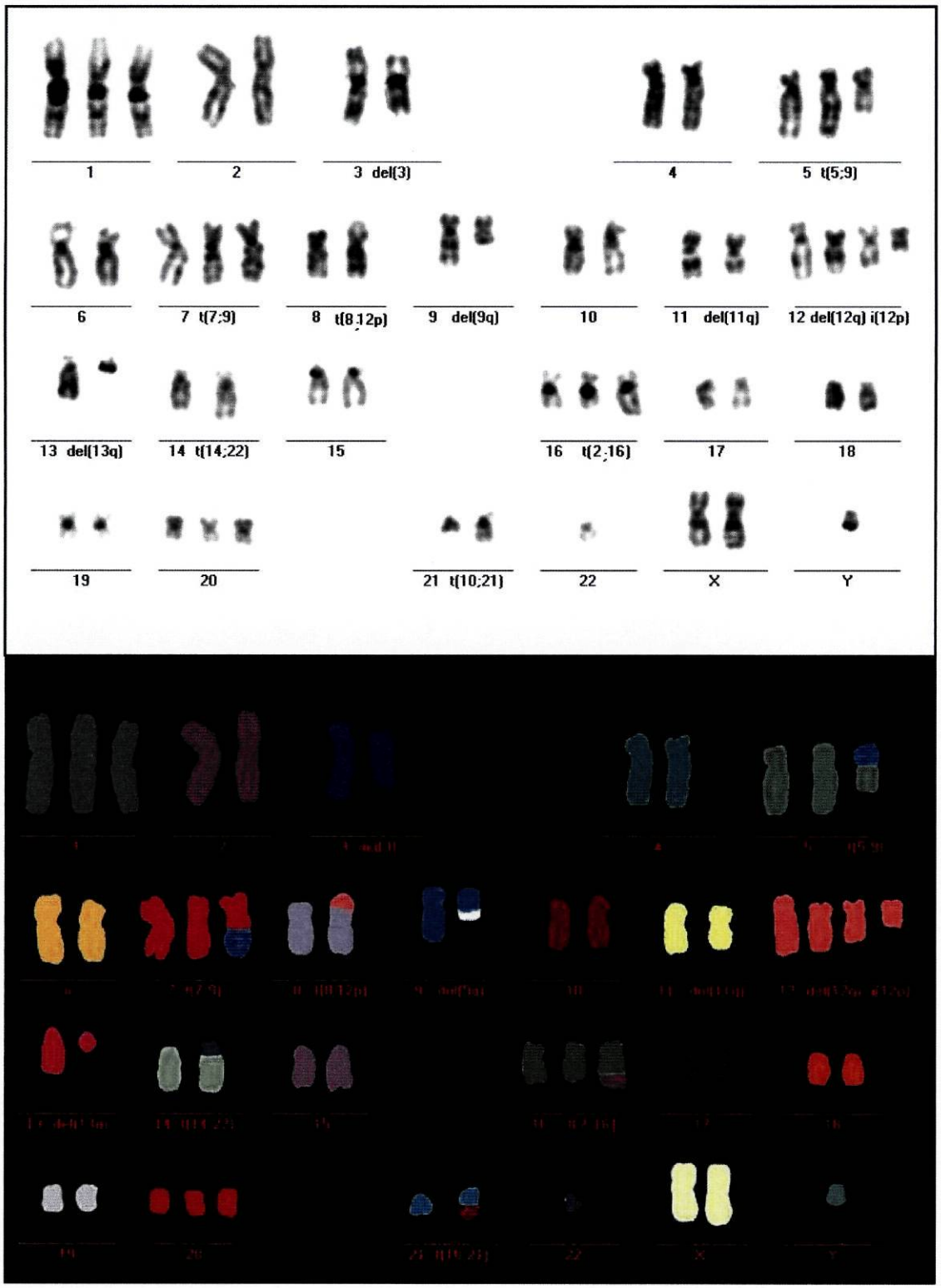
Yoshida M (1995): Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. Bioessays, **17**: 423-430.

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

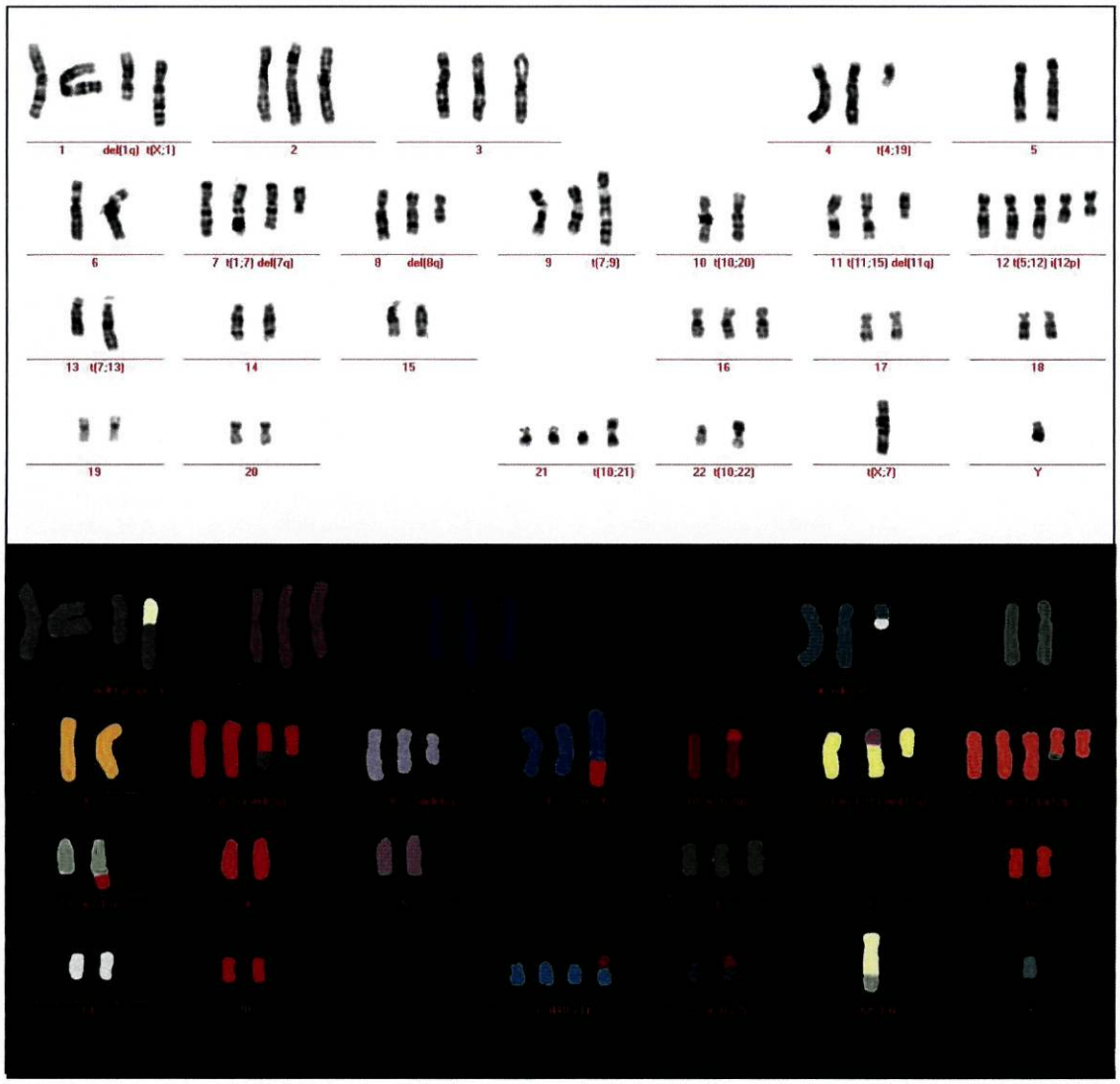
Zhou Z-H, Akgün E, Jasin M (2001): Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. Proc. Natl. Acad. Sci. USA, **98**: 8326-8333.

Appendix

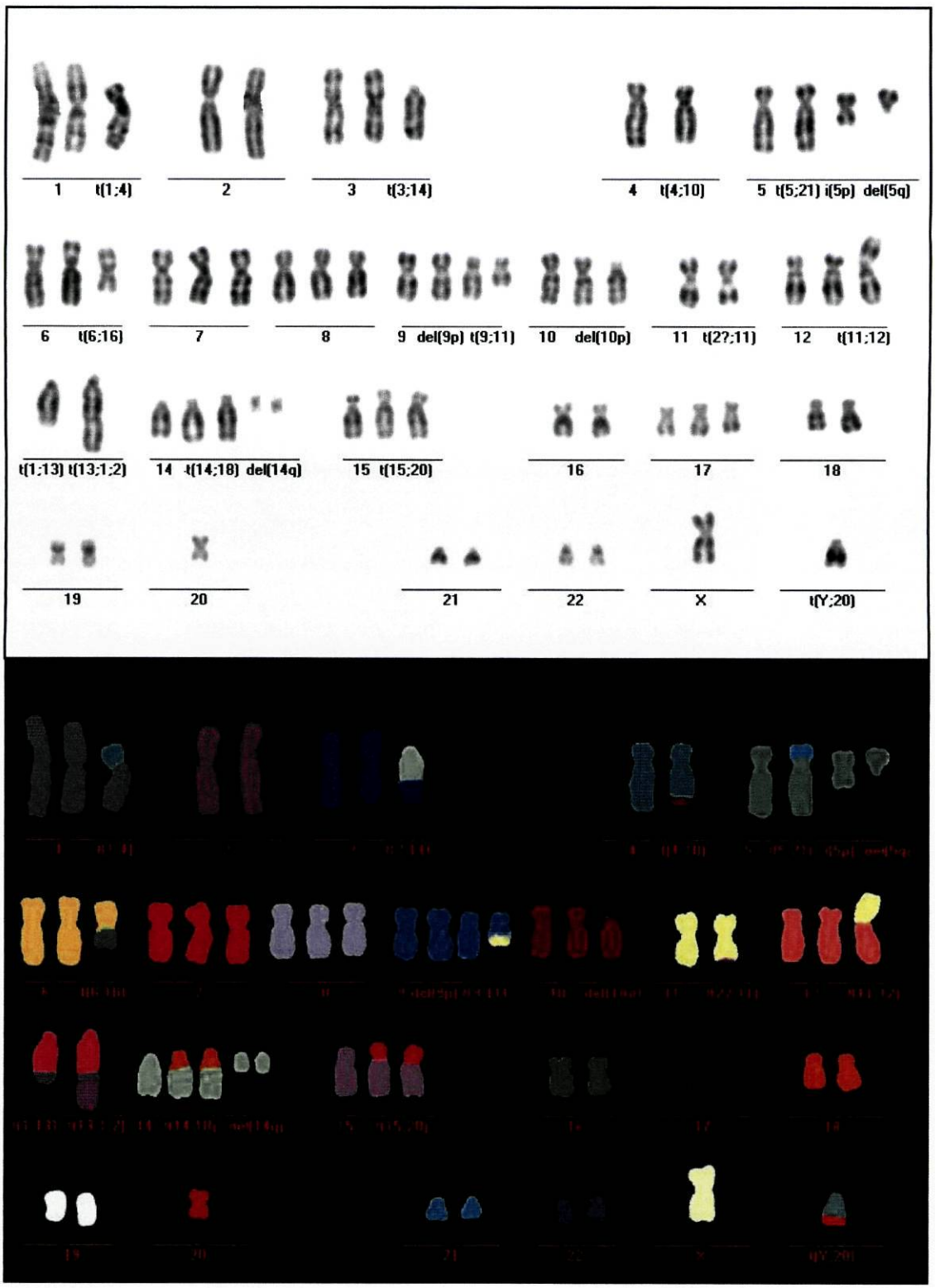
EP 2102


1
2
3 del(3)
4
5 t(5;9)
6
7 t(7;9)
8 t(8;12p)
9 del(9q)
10
11 del(11q)
12 del(12q) i(12p)
13 del(13q)
14 t(14;22)
15
16 t(2;16)
17
18
19
20
21 t(10;21)
22
X
Y

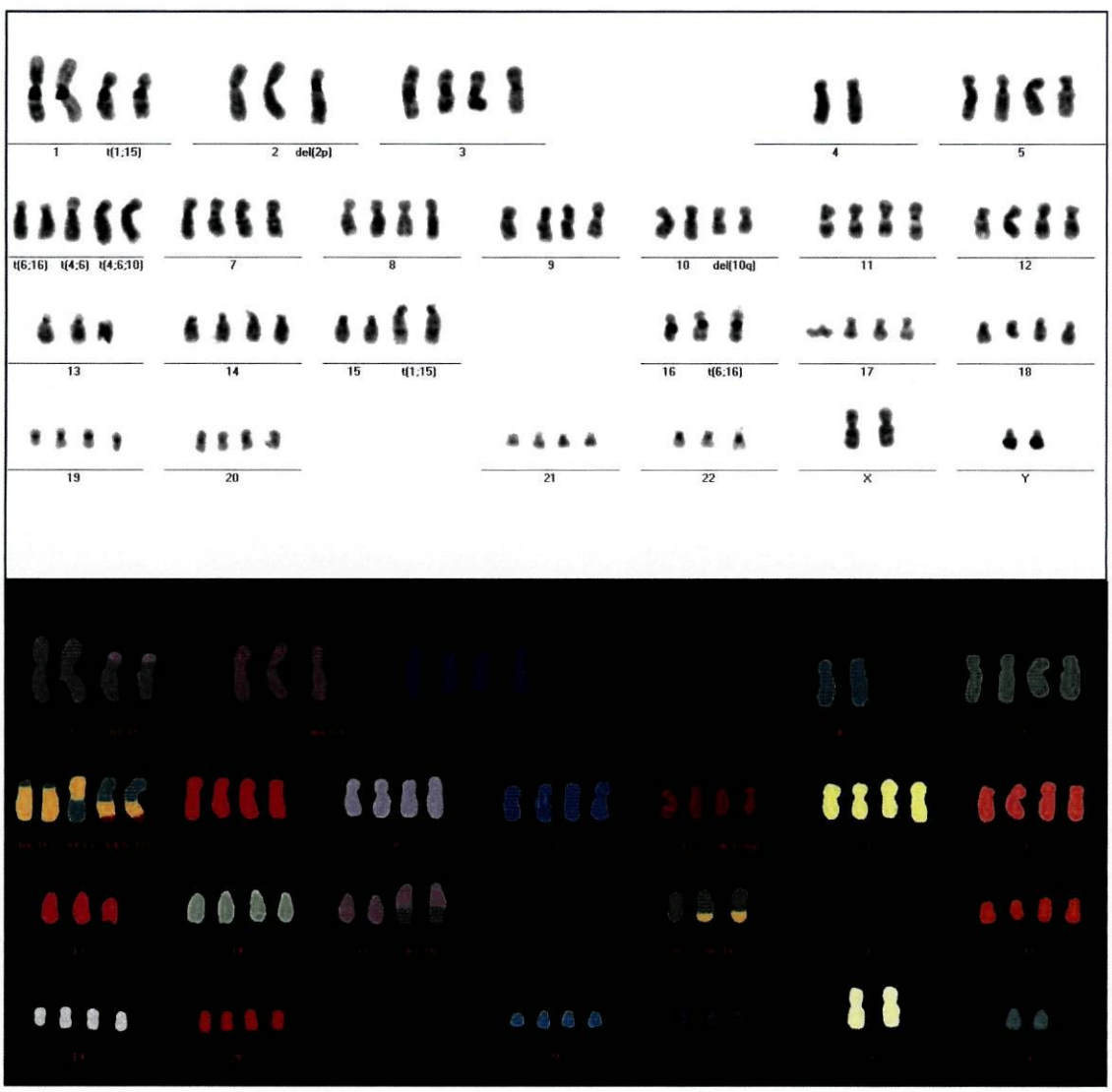
NT2


1 del(1q) t(X;1)
2
3
4 t(4;19)
5
6
7 t(1;7) del(7q)
8 del(8q)
9 t(7;9)
10 t(10;20)
11 t(11;15) del(11q)
12 t(5;12) i(12p)
13 t(7;13)
14
15
16
17
18
19
20
21 t(10;21)
22 t(10;22)
t(X;7)
Y

DU 145


1 t(1;4)
2
3 t(3;14)
4 t(4;10)
5 t(5;21) i(5p) del(5q)
6 t(6;16)
7
8
9 del(9p) t(9;11)
10 del(10p)
11 t(2?;11)
12 t(11;12)
t(1;13) t(13;1;2)
14 t(14;18) del(14q)
15 t(15;20)
16
17
18
19
20
21
22
X
t(Y;20)

LN Cap


1 t(1;15)
2 del(2p)
3
4
5
t(6;16) t(4;6) t(4;6;10)
7
8
9
10 del(10q)
11
12
13
14
15 t(1;15)
16 t(6;16)
17
18
19
20
21
22
X
Y

Papers

ELSEVIER

Molecular and Cellular Endocrinology 166 (2000) 37–43

MCE
Molecular and Cellular Endocrinology

www.elsevier.com/locate/mce

Prognostic factors for successful testicle spermatid recover

M. Sousa [a,*], S. Fernandes [b], A. Barros [b,c]

[a] *Laboratory of Cell Biology, Institute of Biomedical Sciences Abel Salazar, University of Porto, Lg. Prof. Abel Salazar 2, 4099-003 Porto, Portugal*

[b] *Department of Medical Genetics, Faculty of Medicine, University of Porto, Porto, Portugal*

[c] *Centre for Reproductive Genetics A.Barros, Porto, Portugal*

Abstract

Men with non-obstructive azoospermia exhibit different histopathologic syndromes in the testicle biopsy, varying from aplasia, Sertoli cell only syndrome, maturation arrest and hypoplasia. The genetic basis of these syndromes is discussed. We present the diagnostic testicle biopsies performed on 160 consecutive non-obstructive azoospermic patients, and these results were correlated with the findings after multiple bilateral treatment testicle biopsy. Each syndrome had to be reevaluated as for the presence of at least one focus of spermatogenesis up to the primary spermatocyte, round spermatid, elongating spermatid, elongated spermatid, or spermatozoa stages. The clinical outcome using donor sperm-IVF, spermatid or sperm intracytoplasmic injection is thereafter presented. A new prognosis based on the findings of this large clinical series coupled to results obtained with Y chromosome molecular screening is offered. Alternative treatments to donor sperm for men without spermatids are discussed. © 2000 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Azoospermia; Intracytoplasmic injection; Human; Spermatids; Testicle biopsy; Y chromosome

1. Intracytoplasmic injection with sperm from semen, epididymus and testicle

Intracytoplasmic sperm injection was successfully introduced into clinical practice (Palermo et al., 1992; Van Steirteghem et al., 1993a,b) after several unsuccessful trials in humans (reviewed in Palermo et al., 1996). This technique enabled to solve the majority of cases of male infertility due to abnormal concentration, morphology and/or motility of spermatozoa in the ejaculate (Nagy et al., 1995a). The technique could also be successful applied in cases of male immunological infertility (Nagy et al., 1995b), total sperm immotility (Barros et al., 1997a,b), patients with chromosome aberrations (Montag et al., 1997), and anejaculation due to spinal-cord injuries (Barros et al., 1998).

This was later further complemented by sperm extraction from the epididymus in cases of obstructive azoospermia (Tournaye et al., 1994; Silber et al., 1995a), and by sperm extraction from the testicle (TESE) in cases of non-obstructive azoospermic patients (Devroey et al., 1994; Silber et al., 1995a; Devroey et al., 1996; Silber et al., 1996; Nagy et al., 1998). However, some of those azoospermic patients did not present sperm in the testicle tissue, and for these cases no treatment could be found.

2. Intracytoplasmic spermatid injection

Spermatid conception was proposed and clearly anticipated by Edwards et al. (1994), and first successfully achieved by using microinjection of round spermatids (ROSI) retrieved from the semen (three successful cases: Tesarik et al., 1995; Tesarik and Mendoza, 1996; Tesarik et al., 1996; Barak et al., 1998) or testicle (five successful cases: Antinori et al., 1997a,b; Vanderzwalmen et al., 1997; Kaharaman et al., 1998), and of elongating and elongated spermatids (ELSI) recovered from the testicle (27 successful cases: Fishel et al., 1995, 1996; Amer et al., 1997; Antinori et al., 1997a; Araki et al., 1997; Vanderzwalmen et al., 1997; Bernabeu et al., 1998; Kaharaman et al., 1998; Sofikitis et al., 1998; Sousa et al., 1999).

* Corresponding author. Tel.: +351-22-2062217; fax: +351-22-20662232.

E-mail address: msousa@icbas.up.pt (M. Sousa).

0303-7207/00/$ - see front matter © 2000 Elsevier Science Ireland Ltd. All rights reserved.
PII: S0303-7207(00)00295-1

However, large series demonstrated that the use of round spermatids was associated with a very bad prognosis, with pregnancies having been achieved only in cases where the patients had presented mature sperm in previous ejaculates or testicle diagnostic biopsies, thus suggesting that no major genetic causes were present and responsible for the azoospermia situation (Sousa et al., 1998; Tesarik et al., 1998a; Sousa et al., 1999). A bad prognosis has also been found with ELSI using elongating spermatids (Bernabeu et al., 1998; Sousa et al., 1999). Thus, so far, spermatid microinjection works finely only when spermatids appear at the elongated stage and present normally developed flagella. The failure of treatments using the more immature spermatid forms has been attributed to germ cell unrecognized apoptosis, presence of more than one single male germ cell centriole (the normal situation in spermatids), and/or deficiency in sperm-oocyte activating substance (Sousa et al., 1996, 1998; Tesarik et al., 1998a; Sousa et al., 1999). In all these unsuccessful cases, a strong correlation has also been noticed with predominance of unipronuclear development, a finding which suggests that instead of syngamy a single pronucleus may in fact represent parthenogenetic activation (Sousa et al., 1999).

3. Interstitial microdeletions in Yq11

About 11% of idiopatic infertile men present a normal caryotype with microdeletions in Yq11, showing severe oligozoospermia or non-obstructive azoospermia (Vogt, 1996; Vogt et al., 1997; Kostiner et al., 1998; Vogt, 1998).

Comparative molecular screening programs of these patients showed that Yq11 could be resolved in at least 25 subintervals, D1–D25. Patients were then further classified in relation to their specific histopathological testicle syndromes into three main groups: with Sertoli cell only syndrome (microdeletions in D3–D6, locus AZFa, 1Mb length), with maturation arrest up to the primary spermatocyte stage (microdeletions in D13–D16, locus AZFb, 1.5Mb length), and with hypospermatogenesis/severe oligozoospermia (microdeletions in D20–D22, locus AZFc, 3 Mb length; Vogt, 1996; Vogt et al., 1997; Kostiner et al., 1998; Vogt, 1998). Yq11 microdeletions appear more frequent in AZFc, and are due to errors occurring during spermatocyte meiosis of the patients' fathers (hereditary microdeletions), and/or to de novo microdeletions occurring during early embryogenesis (in the majority of the patients). Brakes in the Y chromosome that are responsible for these microdeletions appear to be caused by the presence of repetitive sequences which generate genetic instability at the transition of the three AZF regions, being more numerous in AZFc (Edwards and Bishop, 1997).

Several genes have already been described in all three AZF regions (AZFa, AZFb, AZFc), but their function has not yet been ascribed to any of the testicle syndromes (Vogt, 1996; Vogt et al., 1997; Kostiner et al., 1998; Vogt, 1998).

3.1. Genes in AZFa

In AZFa have been localized the genes DBY (helicases of RNAs metabolism), DFFRY (ubiquitin protease), and UTY (male antigen of the minor histocompatibility H-Y: HYD), with all presenting X homologues (DBX, DFFRX, UTX). Their function remains unknown, and it is supposed that none of them is responsible for the Sertoli cell only syndrome. Due to the homologies with the X chromosome, the AZFa region pairs with the X homologue region during meiosis, which could facilitate microdeletions in this region.

3.2. Genes in AZFb

In AZFb exist several copies of the gene families RBM (D16, D18), and the genes SMCY (D13–D14), TSPY, and E1F1AY.

The locus RBM (RNA Binding Motif) is a gene family whose products are RNA binding proteins. They are testicle specific and do not present X homologues. Only the gene RBM1 (D16) is specifically transcribed in the testicle and appears related to spermatogenesis. The gene RBM1 has at least seven different sequence classes, each class with multiple copies in Yq (with the majority at AZFb) and Yp (associated with copies of the gene TSPY). Patients may present partial or total microdeletions of only one of these classes of the RBM gene family. The gene RBM2 localizes outside of the AZFb region at D18, and is not expressed in the testicle. Genes RBM1/A and RBM1/F are in D15 (AZFb), but it remains unknown if they are functional. Proteins RBM1 have been isolated from the testicle and immunolocalized to the nuclei of spermatogonia, spermatocytes, and round spermatids, but not in elongating or elongated spermatids or in spermatozoa.

Another candidate is gene SMCY (D13–D14 in AZFb). This region also includes the locus Spy, which is related to spermatogenesis, and a family of genes which express the minor histocompatibility antigen H-Y (HLA-B7) and the male antigen SDMA. The gene SMCY has homology with the gene SMCX of the short arm of the X chromosome (Xp11.1-Xp11.2), and both are believed to control sexual differentiation during embryogenesis, with a double dose of Smcx causing female differentiation, and a single dose of each Smcx and Smcy favoring male evolution. Because patients with microdeletions in AZFb and maturation arrest also show deletions in SMCY, in these patients the microdeletion has to have only attained the SMCY

gene after expression of this gene in early embryogenesis, once the male differentiation has not been affected. The locus HY is also supposed to be involved in male differentiation during embryogenesis, before gonadogenesis and spermatogenesis, and is also supposed to be related with locus GBY, present in Yp and Yq11 and responsible for gonadoblastoma. The locus GBY is expressed before spermatogenesis, causing about 30% of the cases of gonadoblastomas in XY gonad dysgenesis (XY females because of SRY deletion in Yp). This can be explained by expression of gene GBY during embryogenesis causing an abnormal growth of female gonadic cells. Expression of gene GBY is also responsible for 55% of gonadoblastoma cases with positive HY antigen, thus suggesting a relationship between the levels of expression of the HY antigen and the tumor developmental potential of gonad remnants.

The gene TSPY is also a family of genes whose majority of copies appear in Yp associated with copies of the RBM gene family, although other copies can also be found in AZFb associated with RBM copies. These genes code for proteins involved in spermatogonia proliferation. The gene E1F1AY was also located to AZFb. It codes for proteins that initiate translation of mRNAs and has an homologue in the X chromosome (E1F1AX).

3.3. Genes in AZFc

Microdeletions in AZFc are associated with azoospermia (with hypospermatogenesis) and severe oligozoospermia. It remains unknown if these two distinct syndromes are due to deletions of different genes or if they represent the natural evolution of a same mutation (oligozoospermia followed by hypospermatogenesis).

The locus DAZ/SPGY belongs to the region AZFc, and contains multiple copies of genes DAZ1 (ex-DAZ) and DAZ2 (ex-SPGY). These genes code for RNA binding proteins, are testicle specific, and have no X chromosome homologues. Their deletion causes severe oligozoospermia and/or non-obstructive azoospermia with hypospermatogenesis. They appear important for the optimal function of the later stages of spermiogenesis but are not involved in the terminal sperm differentiation process itself. This suggests that deletion of these genes do not cause azoospermia but oligozoospermia, which then can evolve with time to azoospermia. The DAZ genes have an homologue in chromosome 3 (DAZL1), which is expressed only in the male germinal tissue. DAZ proteins have been isolated from the testicle, and were immunolocalized to the cytoplasm of elongated spermatids and along the flagellum of mature sperm. This suggests that they transport or store specific mRNAs whose translation appears inhibited up to more advanced stages of the spermiogenetic process.

Several other genes have been found in AZFc (BPY2, PRY, RBM, TTY1, TTY2), each one with multiple copies in that interval. These genes also have copies in the proximal Yp region, but their function in spermatogenesis remains unknown.

3.4. Other genes in Yq11

Several other genes with a potential role in spermatogenesis (BPY1, CDY, TTY1, XKRY, TB4Y) have been identified to Yq11 but outside the AZF regions. They also present copies in Yp. Gene TB4Y has an homologue in the X chromosome; it is not testicle specific, its deletion does not cause infertility, and its function is still to be determined.

3.5. DNA repetitive regions in Yq11

Region AZFc exhibit two polymorphic loci, DYS1 (D22) and DYS2 (D20), which enables the possibility of different patients showing distinct deletions of the same gene. DYS1 has seven sequence variations, and encases DAZ. DYS2 has five sequence variations, with sequence DYS1/C being the most frequently mutated. In some patients with extensive microdeletions, it includes region RBM and DAZ1 of AZFc.

4. Prognostic factors for successful testicle spermatid recover

It is now clear that the histopathological diagnosis after single testicle biopsy, even when performed bilaterally, the testis volume, and the male hormonal profile does not entirely correlate to the findings from multiple bilateral treatment testicle biopsy (Silber et al., 1995b; Tournaye et al., 1996; Silber et al., 1997; Tournaye et al., 1997; Sousa et al., 1999).

We here present a review of our consecutive 160 cycles of multiple bilateral treatment testicle biopsy and Y chromosome molecular screening in patients with male factor infertility due to non-obstructive azoospermia. This is the larger published clinical series, and may enable teams to offer to their patients a prognosis and correctly schedule the treatment approach.

Of the 160 patients analyzed, the diagnostic testicle biopsy have shown three major syndromes, maturation arrest at the primary spermatocyte stage ($n = 70$; 44%), Sertoli cell only syndrome ($n = 55$; 34%), and hypoplasia ($n = 35$; 22%). After multiple bilateral treatment testicle biopsy, about 69% of the maturation arrest cases, about 16% of the Sertoli cell only syndrome cases, and 100% of the hypoplasia cases, showed at least one focus of complete spermatogenesis up to the elongated spermatid or sperm stages, whose cells proved to have full fertilizing potential (Figs. 1–3).

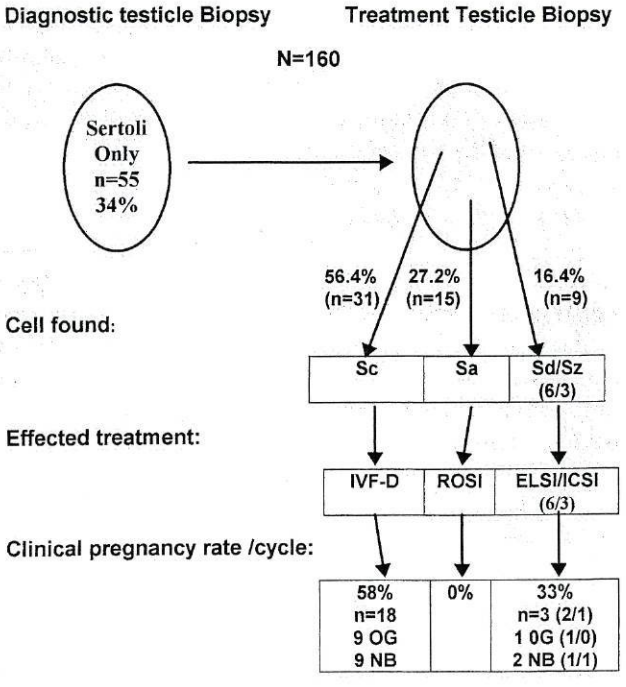


Fig. 1. More mature cells found: Sa, round spermatids; Sc, Sertoli cells; Sd, elongated spermatids; Sz, spermatozoa. Treatments: ELSI, intracytoplasmic elongated spermatid injection; ICSI, intracytoplasmic sperm injection; IVF-D, in vitro fertilization with donor sperm; ROSI, intracytoplasmic round spermatid injection. Pregnancy results: NB, newborns; OG, ongoing pregnancies.

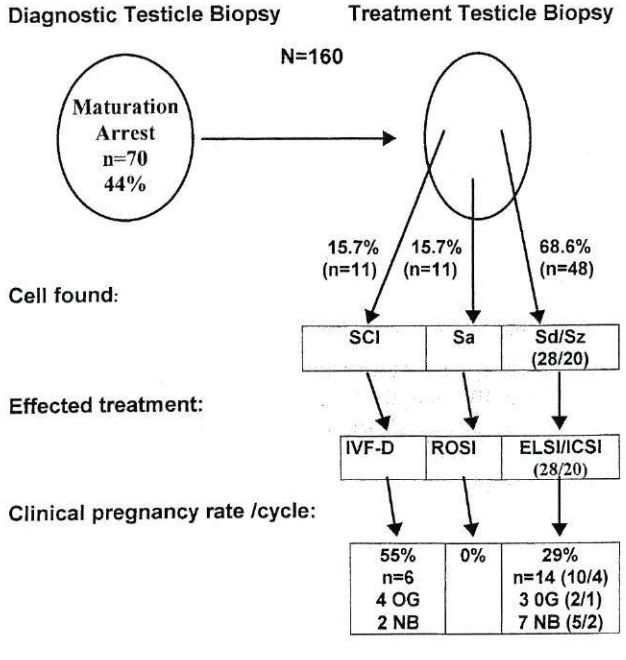


Fig. 2. More mature cells found: Sa, round spermatids; SCI, primary spermatocytes; Sd, elongated spermatids; Sz, spermatozoa. Treatments: ELSI, intracytoplasmic elongated spermatid injection; ICSI, intracytoplasmic sperm injection; IVF-D, in vitro fertilization with donor sperm; ROSI, intracytoplasmic round spermatid injection. Pregnancy results: NB, newborns; OG, ongoing pregnancies.

This shows that the majority of non-obstructive azoospermic patients can be successfully treated despite the bad initial prognosis. These figures are now offered as a general prognosis to patients before initiating the treatment biopsy. In patients with hypoplasia, TESE is performed always after oocyte pick-up; for cases of maturation arrest and Sertoli cell only syndrome, we schedule the treatment biopsy for the day of oocyte pick-up only if patients accept donor sperm-IVF. If patients refuse this approach or donor semen, we prefer to perform the multiple bilateral treatment testicle biopsy without simultaneous female partner controled ovarian hyperstimulation. If a focus of elongated spermatids or spermatozoa is found, the sample is frozen after 2 days of culture to optimize the maturation process (in IVF medium, at 30–32°C, 5% CO_2 in air); if one focus of spermatogenesis up to the spermatid elongating phase is found, the tissue is cultured up to flagella and elongated spermatid forms are found (in IVF medium + rFSH to stimulate spermatogenesis + testosterone to stimulate Sertoli cells and avoid accelerated in vitro cell apoptosis; at 30–32°C, 5% CO_2 in air; Tesarik et al., 1998b,c,d Cremades et al., 1999). In our experience this takes about 2–3 days to attain successful results, with about 16% of the cells evolving to more developed stages in about 42% of the clinical cases. If one focus of spermatogenesis up to the spermatocyte or round spermatid stages is found, we also perform culture of the tissue in the same conditions. However, in these more difficult cases this takes about 7–12 days to attain successful results, with about 6–8% of the cells evolving to elongated spermatids and spermatozoa in about 25% of the clinical cases.

In an effort to improve our prognostic figures, all our studied non-obstructive azoospermic patients were screened for Yq11 microdeletions (Vogt, 1998). Results showed no microdeletions in patients with abnormal caryotypes and hypoplasia, and about 34% of microdeletions in the other cases.

We then correlated Y microdeletions with findings obtained after diagnostic testicle biopsy and the detailed findings obtained with multiple bilateral treatment testicle biopsies. Results suggest that the presence of microdeletions in AZFc can improve the prognosis of our patients, once this deletion was associated with successful recover of elongated spermatids during TESE. However, the improved prognosis appears to be useful only for Sertoli cell only syndrome cases. This is so, because 67% of the patients simultaneously with a Sertoli cell only syndrome diagnosis and deletion in AZFc will have a focus of complete spermatogenesis in TESE, whereas only 40% of the patients simultaneously with a maturation arrest diagnosis and deletion in AZFc will have a focus of complete spermatogenesis in TESE (in opposition to the 69% of success rate given by histopathology). Thus, for Sertoli cell only patients,

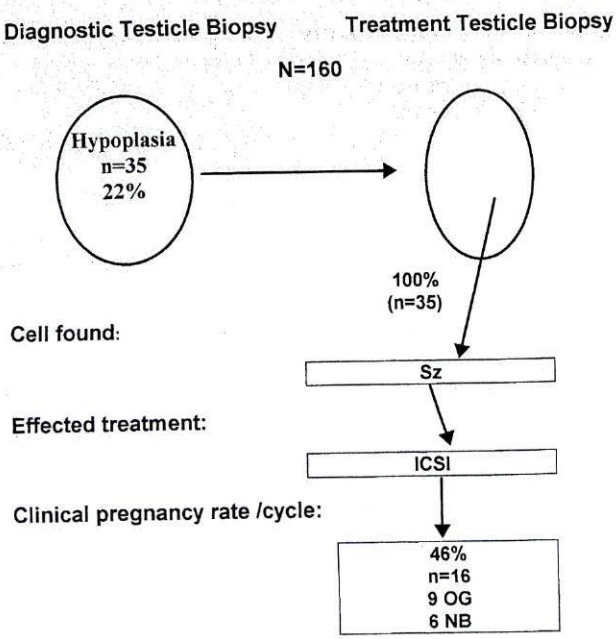


Fig. 3. More mature cells found: Sz, spermatozoa. Treatments: ICSI, intracytoplasmic sperm injection. Pregnancy results: NB, newborns; OG, ongoing pregnancies.

instead of offering a good prognosis of 16% based on histopathology, we can now offer a much better prognosis of 67% if they also present an AZFc mutation.

5. Conclusions and future directions

The large clinical series here presented enables to offer to non-obstructive azoospermic patients a prognosis for successful recover of fertilizing spermatids/spermatozoa during TESE. Patients now program their preferential treatment. In our experience, all patients with hypoplasia, most of the patients with maturation arrest, and all patients with Sertoli cell only syndrome and an AZFc mutation prefer to perform TESE at the time of oocyte recover. Most of our cases with a worse prognosis also prefer the same strategy because they accept to use donor sperm-IVF. On the contrary, patients with a worse prognosis prefer TESE in advance of female partner stimulation when donor sperm is an unacceptable procedure for the couple. Although we offer to all patients tissue culture, the elongated spermatids and spermatozoa produced in vitro are still not used by our team in clinical treatments until our genetic research program is finished. Preliminary results suggest that the normal (2PN, 2PB) fertilization rate (49%), the blastocyst formation rate (22%), and the chromosomal constitution of embryos produced with those in vitro matured cells are still far away from the desirable objectives.

In conclusion, despite these advances, about 43% of the secretory azoospermic population remain without a solution. Although testicle tissue cultures appear promising for about 54% of these cases, it remains urgent to find a new strategy for those men with pure Sertoli cell only syndromes as also for those male patients that have one focus of spermatogenesis up to the spermatocyte stage but whose cells do not mature during in vitro culture.

Acknowledgements

This work was partially funded by public grants from JNICT-PRAXIS-XXI (PCNA/C//BIA/100/96), FCT, and FCT-UMIB.

References

Amer, M., Soliman, E., El-Sadek, M., Mendoza, C., Tesarik, J., 1997. Is complete spermiogenesis failure a good indication for spermatid conception? Lancet 350, 116.

Antinori, S., Versaci, C., Dani, G., Antinori, M., Pozza, D., Selman, H.A., 1997. Fertilization with human testicular spermatids: four successful pregnancies. Hum. Reprod. 12, 286–291.

Antinori, S., Versaci, C., Dani, G., Antinori, M., Selmann, H.A., 1997. Successful fertilization and pregnancy after injection of frozen-thawed round spermatids into human oocytes. Hum. Reprod. 12, 554–556.

Araki, Y., Motoyama, M., Yoshida, A., Kim, S.-Y., Sung, H., Araki, S., 1997. Intracytoplasmic injection with late spermatids: a successful procedure in achieving childbirth for couples in which the male partner suffers from azoospermia due to deficient spermatogenesis. Fertil. Steril. 67, 559–561.

Barak, Y., Kogosowski, A., Goldman, S., Goldman, S., Soffer, Y., Gonen, Y., Tesarik, J., 1998. Pregnancy and birth after transfer of embryos that developed from single-nucleated zygotes obtained by injection of round spermatids into oocytes. Fertil. Steril. 70, 67–70.

Barros, A., Sousa, M., Oliveira, C., Silva, J., Almeida, V., Beires, J., 1997. Pregnancy and birth after intracytoplasmic sperm injection with totally immotile sperm recovered from the ejaculate. Fertil. Steril. 67, 1091–1094.

Barros, A., Sousa, M., Angelopoulos, T., Tesarik, J., 1997. Efficient modification of intracytoplasmic sperm injection technique for cases with total lack of sperm movement. Hum. Reprod. 12, 1227–1229.

Barros, A., Sousa, M., Andrade, M.J., Oliveira, C., Silva, J., Beires, J., 1998. Birth after electroejaculation coupled to intracytoplasmic sperm injection in a gun-shot spinal cord-injured man. Arch. Androl. 41, 5–9.

Bernabeu, R., Cremades, N., Takahashi, K., Sousa, M., 1998. Successful pregnancy after spermatid injection. Hum. Reprod. 13, 1898–1900.

Cremades, N., Bernabeu, R., Barros, A., Sousa, M., 1999. In vitro maturation of round spermatids using coculture on Vero cells. Hum. Reprod. 14, 1287–1293.

Devroey, P., Liu, J., Nagy, Z., Tournaye, H., Silber, S.J., Van Steirteghem, A.C., 1994. Normal fertilization of human oocytes after testicular sperm extraction and intracytoplasmic sperm injection. Fertil. Steril. 62, 639–641.

Devroey, P., Nagy, Z., Tournaye, H., Liu, J., Silber, S., Van Steirteghem, A.C., 1996. Outcome of intracytoplasmic sperm injection with testicular spermatozoa in obstructive and non-obstructive azoospermia. Hum. Reprod. 11, 1015–1018.

Edwards, R.G., Bishop, C.E., 1997. On the origin and frequency of Y chromosome deletions responsible for severe male infertility. Mol. Hum. Reprod. 3, 549–554.

Edwards, R.G., Tarin, J.J., Dean, N., Hirsch, A., Tan, S.L., 1994. Are spermatid injections into human oocytes now mandatory? Hum. Reprod. 9, 2217–2219.

Fishel, S., Green, S., Bishop, M., Thornton, S., Hunter, A., Fleming, S., Al-Hassam, S., 1995. Pregnancy after intracytoplasmic injection of spermatid. Lancet 345, 1641–1642.

Fishel, S., Aslam, I., Tesarik, J., 1996. Spermatid conception: a stage too early, or a time too soon? Hum. Reprod. 11, 1371–1375.

Kaharaman, S., Polat, G., Samli, M., Sozen, E., Ozgun, O.D., Dirican, K., Ozbiçer, T., 1998. Multiple pregnancies obtained by testicular spermatid injection in combination with intracytoplasmic sperm injection. Hum. Reprod. 13, 104–110.

Kostiner, D.R., Turek, P.J., Reijo, R.A., 1998. Male infertility: analysis of the markers and genes on the human Y chromosome. Hum. Reprod. 13, 3032–3038.

Montag, M., Van der Ven, K., Ved, S., Schmutzler, A., Prietl, G., Krebs, D., Peschka, B., Schwanitz, G., Albers, P., Haidl, G., Van der Ven, H., 1997. Success of intracytoplasmic sperm injection in couples with male and/or female chromosome aberrations. Hum. Reprod. 12, 2635–2640.

Nagy, Z.P., Liu, J., Joris, H., Verheyen, G., Tournaye, H., Camus, M., Derde, M.-P., Devroey, P., Van Steirteghem, A.C., 1995. The result of intracytoplasmic sperm injection is not related to any of the three basic sperm parameters. Hum. Reprod. 10, 1123–1129.

Nagy, Z.P., Verheyen, G., Liu, J., Joris, H., Janssenzwillen, C., Wisanto, A., Devroey, P., Van Steirteghem, A.C., 1995. Results of 55 intracytoplasmic sperm injection cycles in the treatment of male immunological infertility. Hum. Reprod. 10, 1775–1780.

Nagy, Z.P., Joris, H., Verheyen, G., Tournaye, H., Devroey, P., Van Steirteghem, A.C., 1998. Correlation between motility of testicular spermatozoa, testicular histology and the outcome of intracytoplasmic sperm injection. Hum. Reprod. 13, 890–895.

Palermo, G., Joris, H., Devroey, P., Van Steirteghem, A.C., 1992. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet 340, 17–18.

Palermo, G.D., Cohen, J., Rosenwaks, Z., 1996. Intracytoplasmic sperm injection: a powerful tool to overcome fertilization failure. Fertil. Steril. 65, 899–908.

Silber, S.J., Nagy, Z., Liu, J., Tournaye, H., Lissens, W., Ferec, C., Liebaers, I., Devroey, P., Steirteghem, A.C., 1995. The use of epididymal and testicular spermatozoa for intracytoplasmic sperm injection: the genetic implications for male infertility. Hum. Reprod. 10, 2031–2043.

Silber, S.J., Steirteghem, A.C., Devroey, P., 1995. Sertoli cell only revisited. Hum. Reprod. 10, 1031–1032.

Silber, S.J., Steirteghem, A.C., Nagy, Z., Liu, J., Tournaye, H., Devroey, P., 1996. Normal pregnancies resulting from testicular sperm extraction and intracytoplasmic sperm injection for azoospermia due to maturation arrest. Fertil. Steril. 66, 110–117.

Silber, S., Nagy, Z., Devroey, P., Tournaye, H., Steirteghem, A.C., 1997. Distribution of spermatogenesis in the testicles of azoospermic men: the presence or absence of spermatids in the testes of men with germinal failure. Hum. Reprod. 12, 2422–2428.

Sofikitis, N.V., Yamamoto, Y., Miyagawa, I., Mekras, G., Mio, Y., Toda, T., Antypas, S., Kawamura, H., Kanakas, N., Antoniou, N., Loutradis, D., Mantzavinos, T., Kalianidis, K., Agapitos, E., 1998. Ooplasmic injection of elongating spermatids for the treatment of non-obstructive azoospermia. Hum. Reprod. 13, 709–714.

Sousa, M., Mendoza, C., Barros, A., Tesarik, J., 1996. Calcium responses of human oocytes after intracytoplasmic injection of leucocytes, spermatocytes and round spermatids. Mol. Hum. Reprod. 2, 853–857.

Sousa, M., Barros, A., Tesarik, J., 1998. Current problems with spermatid conception. Hum. Reprod. 13, 255–258.

Sousa, M., Barros, A., Takahashi, K., Oliveira, C., Silva, J., Tesarik, J., 1999. Clinical efficacy of spermatid conception: analysis using a new spermatid classification scheme. Hum. Reprod. 14, 1279–1286.

Tesarik, J., Mendoza, C., 1996. Spermatid injection into human oocytes. I. Laboratory techniques and special features of zygote development. Hum. Reprod. 11, 772–779.

Tesarik, J., Mendoza, C., Testart, J., 1995. Viable embryos from injection of round spermatids into oocytes. New Engl. J. Med. 333, 525.

Tesarik, J., Rolet, F., Brami, C., Sedbon, E., Thorel, J., Tibi, C., Thebault, A., 1996. Spermatid injection into human oocytes. II. Clinical application in the treatment of infertility due to non-obstructive azoospermia. Hum. Reprod. 11, 780–783.

Tesarik, J., Sousa, M., Greco, E., Mendoza, C., 1998. Spermatids as gametes: indications and limitations. Hum. Reprod. 13 (Suppl. 3), 89–111.

Tesarik, J., Greco, E., Rienzi, L., Ubaldi, F., Guido, M., Cohen-Bacrie, P., Mendoza, C., 1998. Differentiation of spermatogenic cells during in vitro culture of testicular biopsy samples from patients with obstructive azoospermia: effect of recombinant follicle stimulating hormone. Hum. Reprod. 13, 2772–2781.

Tesarik, J., Guido, M., Mendoza, C., Greco, E., 1998. Human spermatogenesis in vitro: respective effects of follicle stimulating hormone and testosterone on meiosis, spermiogenesis, and Sertoli cell apoptosis. J. Clin. Endocrinol. Metabolism. 83, 4467–4473.

Tesarik, J., Bahceci, M., Ozcan, C., Greco, E., Mendoza, C., 1998. Restoration of fertility by in vitro spermatogenesis. Lancet 353, 555–556.

Tournaye, H., Devroey, P., Liu, J., Nagy, Z., Lissens, W., Van Steirteghem, A.C., 1994. Microsurgical epididymal sperm aspiration and intracytoplasmic sperm injection: a new effective approach to infertility as a result of congenital bilateral absence of the vas deferens. Fertil. Steril. 61, 1045–1051.

Tournaye, H., Liu, J., Nagy, P.Z., Camus, M., Goossens, A., Silber, S., Van Steirteghem, A.C., Devroey, P., 1996. Correlation between testicular histology and outcome after intracytoplasmic sperm injection using testicular spermatozoa. Hum. Reprod. 11, 127–132.

Tournaye, H., Verheyen, G., Nagy, P., Ubaldi, F., Goossens, A., Silber, S., Van Steirteghem, A.C., Devroey, P., 1997. Are there any predictive factors for successful testicular sperm recovery in azoospermic patients? Hum. Reprod. 12, 80–86.

Vanderzwalmen, P., Zech, H., Birkenfeld, A., Yemini, M., Bertin, G., Lejeune, B., Nijs, M., Segal, L., Stecher, A., Vandamme, B., van Roosendaal, E., Schoysman, R., 1997. Intracytoplasmic injection of spermatids retrieved from testicular tissue: influence of testicular pathology, type of selected spermatids and oocyte activation. Hum. Reprod. 12, 1203–1213.

Van Steirteghem, A.C., Liu, J., Joris, H., Nagy, Z., Janssenswillen, C., Tournaye, H., Derde, M.P., Van Assche, E., Devroey, P., 1993. Higher success rate by intracytoplasmic sperm injection than by subzonal insemination. Report of a second series of 300 consecutive treatment cycles. Hum. Reprod. 8, 1055–1060.

Van Steirteghem, A.C., Nagy, Z., Joris, H., Liu, J., Staessen, C., Smitz, J., Wisanto, A., Devroey, P., 1993. Higher fertilization and implantation rates after intracytoplasmic sperm injection. Hum. Reprod. 8, 1061–1066.

Vogt, P.H., 1996. Human Y chromosome function in male germ cell development. Adv. Dev. Biol. 4, 191–257.

Vogt, P.H., 1998. Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. Mol. Hum. Reprod. 4, 739–744.

Vogt, P.H., Rappold, G.A., Lau, Y.-F., 1997. Report of the third international workshop on human Y chromosome mapping 1997. Cytogenet. Cell Genet. 79, 1–20.

Molecular Human Reproduction Vol.8, No.3 pp. 286–298, 2002

High frequency of *DAZ1/DAZ2* gene deletions in patients with severe oligozoospermia

S.Fernandes[1,2], K.Huellen[1], J.Goncalves[3], H.Dukal[1], J.Zeisler[1], E.Rajpert De Meyts[4], N.E.Skakkebaek[4], B.Habermann[5], W.Krause[5], M.Sousa[6], A.Barros[2] and P.H.Vogt[1,7]

[1]Reproduction Genetics, Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany, [2]Department of Human Genetics, Faculty of Medicine, University of Porto, Porto, [3]National Institute of Health Dr. Ricardo Jorge, Human Genetics Center, Lisboa, Portugal, [4]Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark, [5]Department of Andrology, University of Marburg, Marburg, Germany and [6]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

[7]To whom correspondence should be addressed at: Im Neuenheimer Feld 328, D-69120 Heidelberg, FRG, Germany. E-mail: peter_vogt@med.uni-heidelberg.de

Deletions of the *DAZ* gene family in distal Yq11 are always associated with deletions of the azoospermia factor c (AZFc) region, which we now estimate extends to 4.94 Mb. Because more Y gene families are located in this chromosomal region, and are expressed like the *DAZ* gene family only in the male germ line, the testicular pathology associated with complete AZFc deletions cannot predict the functional contribution of the *DAZ* gene family to human spermatogenesis. We therefore established a *DAZ* gene copy specific deletion analysis based on the *DAZ*-BAC sequences in GenBank. It includes the deletion analysis of eight *DAZ*-DNA PCR markers [six *DAZ*-single nucleotide varients (SNVs) and two *DAZ*-sequence tag sites (STS)] selected from the 5′ to the 3′end of each *DAZ* gene and a deletion analysis of the gene copy specific *Eco*RV and *Taq*I restriction fragments identified in the internal repetitive *DAZ* gene regions (DYS1 locus). With these diagnostic tools, 63 DNA samples from men with idiopathic oligozoospermia and 107 DNA samples from men with proven fertility were analysed for the presence of the complete *DAZ* gene locus, encompassing the four *DAZ* gene copies. In five oligozoospermic patients, we found a *DAZ*-SNV/STS and DYS1/*Eco*RV and *Taq*I fragment deletion pattern indicative for deletion of the *DAZ1* and *DAZ2* gene copies; one of these deletions could be identified as a 'de-novo' deletion because it was absent in the *DAZ* locus of the patient's father. The same *DAZ* deletions were not found in any of the 107 fertile control samples. We therefore conclude that the deletion of the *DAZ1/DAZ2* gene doublet in five out of our 63 oligozoospermic patients (8%) is responsible for the patients' reduced sperm numbers. It is most likely caused by intrachromosomal recombination events between two long repetitive sequence blocks (AZFc-Rep1) flanking the *DAZ* gene structures.

Key words: AZFc/DAZ deletions/*DAZ*-SNVs/male infertility/oligoazoospermia

Introduction

The *Deleted in AZoospermia* (*DAZ*) gene family is located in the distal euchromatic part of the long arm of the human Y chromosome (Yq11.23) in the so-called azoospermia factor c (AZFc) region (Vogt, 1998; Saxena *et al.*, 2000). It has been transposed to the Y chromosome during primate evolution after the divergence of the New World and Old World monkeys, i.e. 35×10^6 years ago (Shan *et al.*, 1996; Seboun *et al.*, 1997). The ancestor gene *DAZ-Like 1* (*DAZL1*) is located on the short arm of chromosome 3 (3p25). *DAZ* and *DAZL1* are transcribed exclusively in the germ line and encode RNA binding proteins of the highly conserved RNA-recognition motif (RRM) class (Burd *et al.*, 1996).

The *DAZL1* gene is an essential master gene for the pre-meiotic development of male and female germ cells (Ruggiu *et al.*, 1997; Houston *et al.*, 1998). The function of the *DAZ* genes in germ cells is not yet known. Human DAZ proteins have been found in late spermatids and in sperm tails (Habermann *et al.*, 1998), and in spermatogonia and spermatocytes (Reijo *et al.*, 2000), and a *DAZ* transgene can partially rescue the pathological phenotype of a *DAZL1* knock-out mutant in the mouse (Slee *et al.*, 1999). This suggests that the *DAZ* genes are functional at different phases of human spermatogenesis and might have evolved to improve or extend specifically the male germ line function of *DAZL1* in Old World monkeys. However, deletion of the complete *DAZ* gene family in humans is compatible with the formation of mature spermatozoa, although only in small numbers; severe oligozoo-

© European Society of Human Reproduction and Embryology

spermia is the primary phenotype (Vogt, 1998) and rare cases exist where complete *DAZ* deletions have been transmitted from father to son (Vogt *et al.*, 1996; Kleiman *et al.*, 1999). Comparative nucleotide mutation rates in *DAZ* exon and intron sequences have also led to the assumption that the Y-located *DAZ* gene family plays only a limited role in human spermatogenesis (Agulnik *et al.*, 1998).

An association between specific *DAZ* gene mutations and a particular testicular pathology is still lacking, because all AZFc deletions described so far remove not only all of the *DAZ* genes, but also all other Y gene families which have been mapped to the AZFc interval (Lahn and Page, 1997; Vogt, 1998). Like the *DAZ* gene family, all these genes are expressed in testicular tissue only and therefore are also AZFc candidate genes.

The number and structure of the *DAZ* gene copies on the human Y chromosome is assumed to be variable in different men, ranging between three and seven gene copies (Yen *et al.*, 1997; Gläser *et al.*, 1998; Saxena *et al.*, 2000) and the structures of the *DAZ* genes in the GenBank BAC sequences (Tilford *et al.*, 2001) have been proposed to be similar to those present in the *DAZ*-BACs isolated by Saxena *et al.* (Saxena *et al.*, 2000). However, their detailed structures are not yet described. We assumed that these structures may be different because the BACs from GenBank were cloned from a different male donor (RPCI-11-donor) (Osoegawa *et al.*, 1998) than the one used to clone the BACs isolated by Saxena *et al.* (CTA/CTB-donor) (Saxena *et al.*, 2000; Shizuya *et al.*, 1992). We therefore assembled a contig of 44 RPCI-11-BAC sequences from the GenBank data base, covering the complete AZFc region (4.94Mb, AZFc-BAC contig) and analysed the structures of the *DAZ* genes in this contig in detail. Only the *DAZ1* gene structure was found to be identical to that proposed by Saxena *et al.* in the CTA/CTB BAC sequences (Saxena *et al.*, 2000). Substantial differences were found in the repetitive exon regions of the *DAZ2*, *DAZ3* and *DAZ4* genes structures. In the AZFc RPCI-11-BAC contig, the molecular distance between the two identified *DAZ* gene doublets was estimated to be 1.47 Mb.

Based on the *DAZ* gene sequences of the RPCI-11-*DAZ*-BACs, we established a gene copy specific marker analysis with six *DAZ*-single nucleotide variants (SNVs) and two *DAZ*-sequence tagged sites (STSs). Presence or absence of a specific *DAZ* gene copy could be correlated with the presence or absence of a specific set of SNV alleles and/or STS markers (*DAZ* deletion haplotype). This SNV/STS deletion analysis was confirmed by a deletion analysis of *DAZ* gene copy specific *Eco*RV and *Taq*I restriction fragments of the DYS1 locus (Ngo *et al.*, 1986) in DNA blot experiments. The DYS1 DNA locus includes the repetitive *DAZ* exon 7 regions (Vogt *et al.*, 1997).

We used this analysis to investigate the presence of putative *DAZ* gene deletions in the Y chromosomes of 63 infertile men with idiopathic oligozoospermia and of 107 men with normal fertility. In five oligozoospermic patients, we found *DAZ*-SNV/STS and DYS1-*Eco*RV/*Taq*I fragment deletion patterns indicative for deletion of the *DAZ1* and *DAZ2* gene copies, one of which could be identified as a 'de-novo' deletion event because it was absent in the *DAZ* locus of the patient's father. Since the same *DAZ* deletion haplotype was not found in any of the 107 fertile control samples, we assume that the deletions of the *DAZ1/DAZ2* gene doublet in five of our 63 oligozoospermic patients (8%) is the causative agent for the patients' reduced sperm numbers.

Materials and methods

Selection of oligozoospermic patients and the fertile men group

We first selected 77 infertile patients with a clinical diagnosis of idiopathic oligozoospermia, i.e. gonadal epididymal and prostatic abnormalities and testis tumours were excluded in all patients. Their sperm numbers ranged between 0.1 and 10×10^6 per ml of seminal fluid. Sperm motility and morphology were found to be below the World Health Organization reference values (World Health Organization, 1999) for spermatozoa with normal fertilization capacity. All patients were also analysed for the presence of a normal karyotype, 46XY, and for the absence of Yq11 microdeletions using the STS marker set which we established previously for the detection of complete AZFa, AZFb and AZFc deletions (Vogt *et al.*, 1996). Using gene-specific STSs we also analysed each patient for a possible deletion of other AZFc candidate genes. According to these criteria, 63 of the 77 patients were found to be 'normal'. Only these individuals were included in our *DAZ* gene copy specific deletion analysis (OZ samples).

The fertility status of 107 fertile men (designated as 'FM samples') was proven by the fact that they had fathered one or more children. In addition, the normal reproductive status of 51 Danish FM controls (italic numbers in Table III) was confirmed by a normal sperm count ($>20\times10^6$ per ml semen fluid), normal motility and morphology and normal levels of inhibin B, testosterone, LH and FSH according to the guidelines of the World Health Organization.

SNVs/STSs PCR assays for partial DAZ deletions

Of the six *DAZ*-SNVs and two *DAZ*-STS listed in Table I, three SNVs (I, IV, VI) and the two STSs specific for the detection of the *DAZ1* and *DAZ4* genes (*DAZ*-RRM3) and the *DAZ3* gene (*Y-DAZ3*) were novel. Their primer sequences and respective PCR conditions were deposited in Genbank (accession numbers: Table I). The PCR reactions were performed in a 25 µl reaction volume with 1×PCR buffer (Gibco-BRL, Life Technologies GmbH, Karlsruhe, Germany), 2.5 mmol/l $MgCl_2$ (SNV III, IV, Y-*DAZ* 3) or 2.0 mmol/l $MgCl_2$ (SNV I, II, V, VI, *DAZ*-RRM3), 0.2 mmol/l dNTPs, 1 IU *Taq* DNA polymerase (Gibco) and 100 ng of genomic DNA. For the STS analyses, we developed multiplex PCR assays, containing two internal positive controls, one for the SRY locus (sY14) and one to confirm the proposed *DAZ1/DAZ4* deletions (sY152), analysed also by *DAZ*-RRM3 STS. Genomic DNA samples from a fertile male and a female were used as positive and negative controls respectively in each PCR experiment. The DNA sample from the fertile man controls the sensitivity and specificity of each performed PCR assay, whilst the female DNA sample controls the Y specificity of the used primer pairs. In addition, a water sample containing all components except genomic DNA was run with each set of primers in order to detect any possible reagent contamination.

The PCR reaction profile (for Biometra *T*-Gradient Personal Cycler) was as follows: 35 PCR cycles with a pre-soak for 3 min at 94°C, denaturation for 1 min at 94°C, annealing for 1 min at the annealing temperature specific for each primer pair, with a subsequent polymerization step of 1 min at 72°C and, at the end, a final extension step of 3 min at 72°C. The PCR products (5 µl aliquots for SNVs and 15 µl

Table I. *DAZ*-SNVs and STSs, with different alleles in *DAZ1-DAZ4* gene copies

DAZ-SNVs	*DAZ* gene region	GenBank accession no.	SNV restriction enzyme	*DAZ* gene copy[a]	SNV allele[a]	Restricted fragments (bp)
I	5′ UTR-exon1	G73167	*Fsp* I	1,2,3	A	709
				4	**B**	398+311
II	exon 3–exon 4	G73163	*Mbo* I	**1**	**A**	182
				2,3,4	B	122+60
III (sY586)	intron 7A/B	G63907	*Taq* I	**2**	**A**	301
				1,3,4	B	184+117
IV	intron 7H	G73168	*Alu* I	**2**	**A**	630
				1,3,4	B	398+262
V (sY587)	intron 9	G63908	*Dra* I	3,4	A	195+49
				1,2	B	122+73+49
VI	3′ *DAZ*	G73169	*Afl* III	1,2,3	A	431
				4	**B**	248+183
DAZ locus STSs	*DAZ* gene region	GenBank accession no.	BAC clone(s) in *DAZ* locus contig	Presence in *DAZ* copy	Anneal. temp. (°C)	PCR product (bp)
DAZ-RRM3	*DAZ1+4* intron 6b[b]	G73171	BAC140H23; AC053490 BAC539D10; AC006338	1+4	66	473
Y-*DAZ3*	3′ *DAZ*	G73170	BAC290O03; AC010089	**3**	66	260
DAZ1+4 (sY152)	*DAZ1+4* intron 6a[b]	G12002	BAC140H23; AC053490 BAC539D10; AC006338	1+4	66	125
sY14	*SRY*	G38356	–	Control	66	470

[a]STS/SNV alleles in bold letters and numbers mark their specificity for one *DAZ* gene copy.
[b]Intron 6a and 6b are proximal (a) and distal (b) to exon 7G only present in 10.8 kb repeat in *DAZ1* and *DAZ4* (Figure 1B).

aliquots for STSs) were analysed on agarose gels stained with ethidium bromide. GeneRuler^{TM} 100 bp DNA ladder-plus (Gibco) was used as molecular size marker. A positive result for STS markers was indicated by the presence of the respective fragment on the gel. For the subsequent SNVs analyses, the restriction digestions with the enzymes listed on Table I were performed in a 25 µl reaction volume with the appropriate 1× buffer (MBI-Fermentas GmbH, St Leon-Rot, Germany for SNV II, III, IV, V and New England Biolabs GmbH, Frankfurt, Germany for SNV I and VI), 10 units of each enzyme and 20 µl of the PCR solution. Incubations were for 2 h at 37°C, except for *Taq*I digestion which was performed at 65°C. The digests (25 µl aliquots) were analysed on agarose gels (2 or 3% depending on the fragment size obtained for each SNV; Table I) and stained with ethidium bromide (see Figure 2).

Southern blot analysis

A total of 8 µg of purified genomic DNA were digested with 2.5 IU/µg of the restriction enzyme *Eco*RV or *Taq*I (MBI). The fragments were separated by electrophoresis on 0.8% agarose gels (LE-SeaKem; FMC® Bioproducts, Rockland, Maine, USA) and transferred to nylon membranes (Hybond-N$^+$; Amersham Pharmacia, Freiburg, Germany) by vacuum or capillary blot. Filters were UV-cross-linked (1200 J) and baked at 80°C for 2 h. These were pre-hybridized for 2 h and hybridized overnight at 65°C in Church buffer [7% sodium dodecyl suphate (SDS), 0.5 mol/l $NaHPO_4$, 1 mmol/l EDTA, pH7.2] in roller bottles. The probe used was the 2.8 kb *Eco*RI fragment of plasmid p49f, gel-purified and labelled with ^{32}P dCTP by random priming. After overnight hybridization, filters were washed for 30 min at 65°C in 2×standard saline citrate/0.1% SDS and exposed for 1–6 days at –70°C with Kodak XAR films and intensifying screens (REGO) for visualization of the 49f cross-hybridizing genomic DNA fragments by autoradiography.

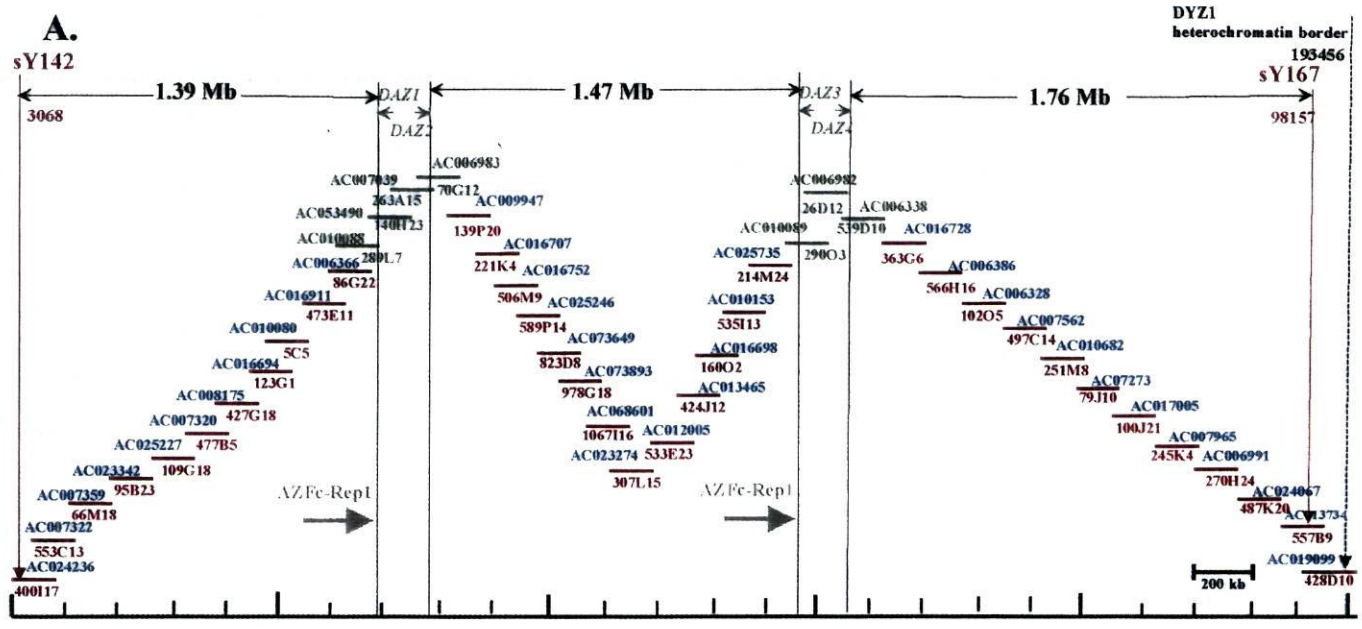
Sequence analysis

For sequence homology analysis in the National Center for Biotechnology Information (NCBI) and GenBank databases, restriction mapping and multiple alignment analysis, we used the HUSAR software package of the German Cancer Research Centre, Heidelberg (http://genome.dkfz-heidelberg.de/biounit). Database searches with the BLAST programs of the NCBI (http://www.ncbi.nlm.nih.gov) for the identification of the exon–intron structures of each *DAZ* gene copy were performed with a complete cDNA sequence of the *DAZ2* gene (CT351Y, Genbank AF414184) using the universal splicing rules. The AZFc BAC contig was established by identifying first the BAC clones containing the AZFc border STSs: sY142 (BAC400I17) and sY167 (BAC557B9). Then inspection of the overlap data provided by GenBank were used to collect the internal 41 AZFc RPCI-11 BACs and assemble them in the unique order shown in Figure 1A. Only BACs with certified sequences (i.e. all regions were covered by sequences from more than one subclone and the assembly was confirmed by restriction digest) were selected. The sequence assembly of the RPCI-11-*DAZ* BAC contigs shown in Figure 1B was used to construct the restriction maps of *Eco*RV, *Taq*I and *Mlu*I in the genomic *DAZ* gene structures.

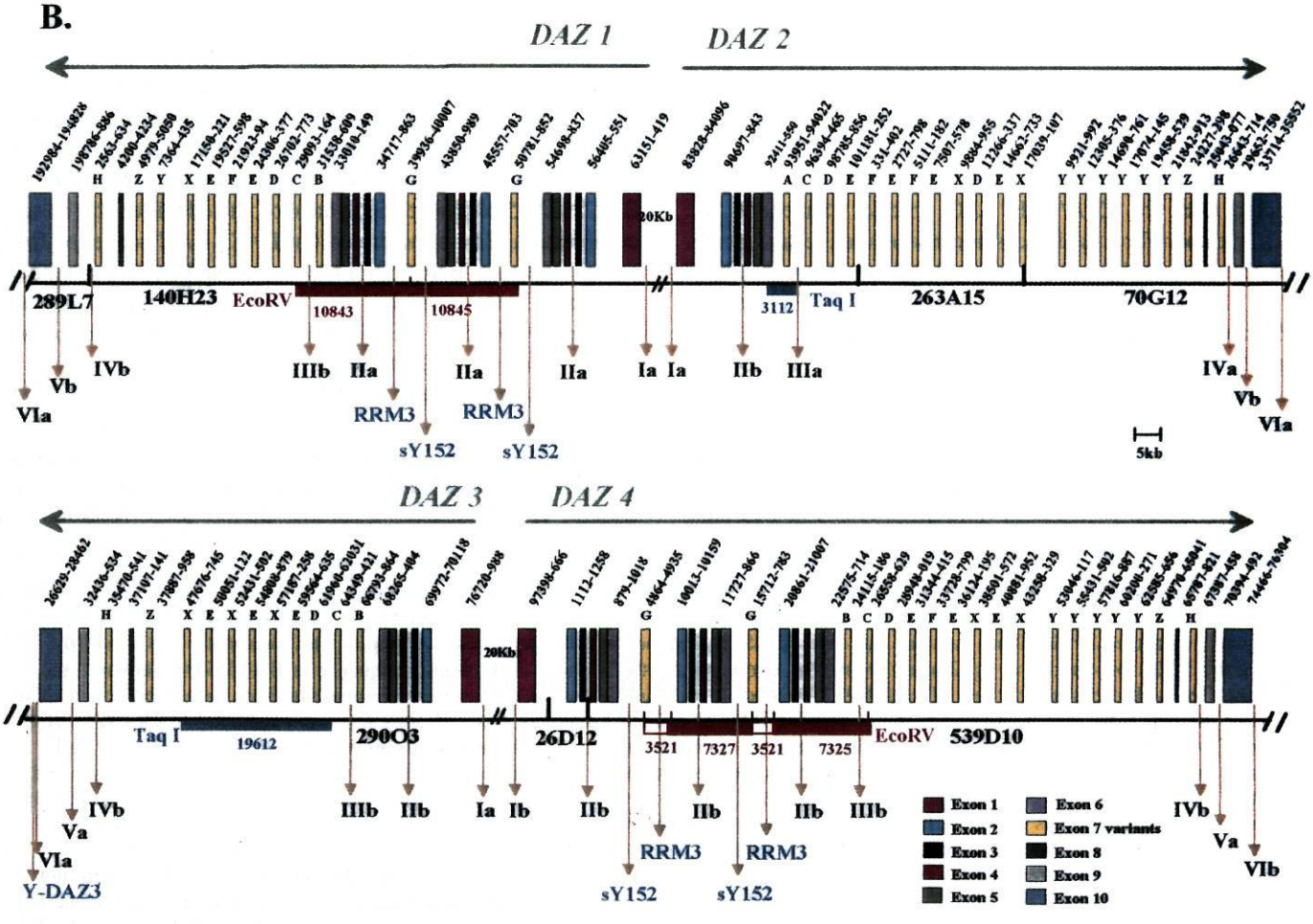
Results

Mapping and structure of four DAZ gene copies in the RPCI-11-AZFc BAC contig

For the development of a *DAZ* gene copy specific deletion analysis, the knowledge of the detailed structure of each *DAZ* gene copy in the RPCI-11-BAC sequences was an essential prerequisite. We therefore screened the GenBank database with the complete *DAZ* cDNA sequence CT351Y (GenBank



AZFc region →Yq11 sequence length between STS sY142 and sY167→ 4.94 Mb



accession no. AF414184) to identify all *DAZ*-RPCI-11-BAC clones. Seven *DAZ*-BACs were identified: 26D12 (163948 bp; GenBank accession no. AC006982); 70G12 (180345 bp; GenBank accession no. AC006983); 140H23 (103432 bp; GenBank accession no. AC053490), 263A15 (168639 bp; GenBank accession no. AC007039); 289L7 (199458 bp; GenBank accession no. AC010088); 290O3 (103356 bp; GenBank accession no. AC010089) and 539D10 (186838 bp; GenBank accession no. AC006338). They were assembled in two small contigs of the large AZFc contig which we constructed by a successive analysis of the GenBank sequence overlap data of 44 RPCI-11 BAC sequences (i.e. 4.94 Mb; Figure 1A). The AZFc contig starts with the BAC400I17 sequence, containing the STSs sY142 (DYS230, GenBank accession no. G38345; sequence position: 3068–3268) and sY143 (DYS231; GenBank accession no. G38347; sequence position: 3074–3506). Both STSs have been identified as markers of the proximal AZFc border line (Simoni *et al.*, 1999). The contig ends with the BAC557B9 sequence, which contains STS sY167 (DYS248; GenBank accession no. G12008; sequence position: 98016–98132). This STS is present in the Y chromosome of all our patients with an AZFc deletion (43 individuals) and therefore was designated as a marker for the distal AZFc border line (S.Fernandes and K.Huellen, unpublished results). Within this AZFc sequence, 1.39 Mb distal to sY142, the *DAZ*-contig containing BACs 289L7, 140H23, 263A15 and 70G12 was mapped proximal to the *DAZ*-contig including BACs 290O3, 26D12, 539D10 with a molecular distance of ~1.47 Mb between them (Figure 1A).

The proximal *DAZ*-BAC contig contains two *DAZ* gene copies with a 3′–5′:: 5′–3′ head-to-head orientation, a similar structure as found earlier in one of the *DAZ*-BAC contigs extracted from the CTA/CTB BAC libraries (Saxena *et al.*, 2000). We therefore used the same nomenclature and designated them as *DAZ1* starting with its 3′ end in BAC289L7 and *DAZ2* ending with its 3′ end in BAC70G12 (Figure 1B). *DAZ1* contains three tandem repeats of 10.8 kb, including exons 2–6 and a tandem repetitive structure of nine exon 7 copies. Nine sequence variants of exon 7 have been found in testicular *DAZ* cDNAs (Vogt *et al.*, 1997; Yen *et al.*, 1997). According to the nomenclature of Yen *et al.* we named them *A*, *B*, *C*, *D*, *E*, *F*, *X*, *Y*, *Z* with the *Y* variant as the most prominent consensus sequence (Yen *et al.*, 1997). Eight variants are found in *DAZ1* and organised in the order: 5′-*B-C-D-E-F-E-X-Y-Z*-3′ (Figure 1B). Since identical repetitive exon 7 structures linked to a triplicated *DAZ* exon 2–6 domain have been found in testicular *DAZ* cDNAs (Saxena *et al.*, 2000), we can assume that the *DAZ1* gene is transcribed in human testis. Two additional exon 7 sequence variants with normal splicing signals were found in intron 6 of the first and second 10.8 kb repeat (exon 7*G* variant; in *DAZ1*: BAC140H23 sequence position: 39936–40007 and 50781–50852) and one in the intron 8 sequence (exon 7*H* variant; in *DAZ1*: BAC140H23 sequence position: 2563–2634) of the genomic *DAZ* gene structures (Figure 1B). However, no *DAZ* cDNA sequence from human testis tissue has yet been found to include one or both of these *DAZ* exon 7 variants. *DAZ1* transcripts spliced with these variants must therefore be rare or not found in testis tissue.

The *DAZ2* gene structure maps to three overlapping BAC sequences. Typical for *DAZ2* is the presence of the *A* variant and absence of the *B* variant in the repetitive exon 7 structure next to six tandem *Y* variants (Figure 1B). *DAZ2* transcripts with this exon 7 variant structure have been found in testicular *DAZ* cDNA sequences (Yen *et al.*, 1997; Saxena *et al.*, 2000), our cDNA screening probe CT351Y being one example (GenBank accession no. AF414184). This indicates that the *DAZ2* gene is also transcribed in human testis tissue. However, there have been no reports of *DAZ2* testis transcripts containing a duplicated *E–F* variant motif and an exon 7*D* variant flanked proximal by exon 7*X* or distal by exon 7*E*, as would be predicted by the genomic *DAZ2* sequence in BAC263A15 (Figure 1B). Such *DAZ2* transcripts might therefore be rare or do not exist. Alternatively, an artifact of the sequence assembly programmes used for the construction of the genomic BAC263A15 sequence might have produced this exon 7 variant composition. The structure of the *DAZ2*-BAC sequence from the CTB library (BAC352E14) contains an *A* and seven tandem *Y* variants in the repetitive exon 7 structure, but the duplicated *E–F* motif found in the RPCI-11 BAC263A15 sequence was absent (Saxena *et al.*, 2000).

In the distal region of the AZFc contig (Figure 1A), 1.47 Mb distal to the 3′end of *DAZ2*, the sequence of BAC290O3 contains the third *DAZ* gene copy, followed by the overlapping BACs 26D12 and 539D10 with the fourth *DAZ* gene copy (Figure 1B). Both have a similiar 3′–5′:: 5′–3′ head-to-head

Figure 1. (**A**) Schematic view of the AZFc RPCI-11 BAC contig between the STS loci sY142 and sY167 (i.e. 4.94 Mb). Their continuous overlap was guaranteed by GenBank data and our own BLAST2 alignments of the proposed overlapping end fragments. BAC428D10, including the DYZ1 border of the Yq12 heterochromatin, overlapped with the distal AZFc border sequence in BAC557B9 (shown at the right). The GenBank accession number of each BAC sequence (blue colour) is given above the corresponding BAC clone (red colour). The location of the two *DAZ* gene doublets and the underlying BAC clones are marked in green. A long repetitive sequence block (AZFc-Rep1, grey arrow) with the same polarity was identified proximal to *DAZ1* and *DAZ3*. The molecular distances between the STSs sY142 and sY167 and the *DAZ1/DAZ2* and *DAZ3/DAZ4* gene doublets were estimated by summarizing the length of all BACs spanning the gap and subtraction of their overlaps. (**B**) Detailed view of the two *DAZ*-BAC contigs for inspection of their different exon structures described in detail in the text. The numbers above each exon (see colour code on the right) correspond to their extension in the underlying BAC sequence. For clarity, these numbers were omitted for the densely spaced exon 3–5 regions. The letters above the exon 7 variants correspond to the nomenclature given by Yen *et al.* (Yen *et al.*, 1997) with the exception of the *G* and *H* variants. These exons were not found in testicular *DAZ* transcripts. The locations and extensions of the DYS1 *Eco*RV and *Taq*I fragments identified as specific for one *DAZ* gene copy (see text) are drawn in red and blue respectively, below each contig. The contigs are drawn to scale with a 20 kb gap in the *DAZ* promoter regions. The locations of STSs, sY152 and (*DAZ*)-RRM3 in *DAZ1* and *DAZ4*, Y-*DAZ3* in *DAZ3*, and the A and B alleles of the six *DAZ*- SNVs (I-VI a/b) in each *DAZ* copy are marked with orange arrows.

orientation as found for the *DAZ3* and *DAZ4* gene structures extracted from the CTB-BAC library (Saxena *et al.*, 2000). However, substantial differences were found in their repetitive exon structures.

Typical for *DAZ3* in the CTB-BAC132B16 sequence was the absence of the *Y* variants and a triplication of the *E–X* motif in the repetitive exon 7 structure (Saxena *et al.*, 2000). This was also found in the *DAZ3* gene of the RPCI-11-BAC290O3 sequence (Figure 1B). However, two additional exon 7 variants (*E* and *F*) in the CTB-*DAZ3* gene were not present in *DAZ3* of the RPCI-11-BAC. Since testicular *DAZ* cDNA sequences with both exon 7 variant compositions have been found (Yen *et al.*, 1997), we can assume that the repetitive sequence region of the *DAZ3* gene structure can be variable on different Y chromosomes, as marked here by the presence or absence of the exon 7 *E–F* motif.

The *DAZ4* gene structure in the RPCI-11 BAC539D10 sequence is substantially different from that of *DAZ4* in the CTB-BAC546E5 sequence. Both *DAZ4* genes were recognised by their 10.8 kb repeat unit; however, three copies of this large repeat unit were found in the RPCI-11-*DAZ4* gene and only two 10.8 kb repeats were found in the CTB-*DAZ4* gene (Saxena *et al.*, 2000). The number and composition of the exon 7 variants in both *DAZ4* genes was also different. We found 15 exon 7 copies with the composition 5′-*B-C-D-E-F-E-X-E-X-Y-Y-Y-Y-Y-Z*-3′ in the RPCI-11-*DAZ4* gene, whereas only nine exon 7 repeats with the composition 5′-*B-C-D-E-F-E-X-Y-Z*-3′ were identified in the *DAZ4* gene of the CTB-BACs (Saxena *et al.*, 2000). Testicular *DAZ* cDNAs with an exon 7 variant structure typical for the RPCI-11-*DAZ4* structure (i.e. *B* variant linked to five tandem *Y* variants) have been described by Yen *et al.* (Yen *et al.*, 1997), although complete *DAZ4*-cDNAs with 15 exon 7 copies, linked to a triplicated 10.8 repeat, have not yet been isolated.

Establishment of a DAZ gene copy specific deletion analysis

DAZ gene copy specific sequence variants (SNVs) and STSs *DAZ* gene locus restriction maps and sequence analyses have shown that the genomic structures of the four *DAZ* genes present in the AZFc region are more than 99% homologous to each other (Saxena *et al.*, 1996, 2000; J.Zeisler and S.Fernandes, unpublished results). For the identification of *DAZ* gene copy specific deletion events, our search for *DAZ* copy specific markers along the complete gene sequences therefore first focused on a systematic identification of small sequence variants, including specific restriction sites (*DAZ*-SNVs) and copy-specific *DAZ*-STSs in the RPCI-11-*DAZ*-BAC sequences. We established six *DAZ*-SNVs (I–VI) spanning the 5′ to the 3′ end of these *DAZ* gene sequences and two *DAZ* gene copy-specific STSs (*DAZ*-RRM3, Y-*DAZ3*; Table I). We would like to acknowledge the work of our colleagues who first identified the SNV in intron 3 of the *DAZ1* gene (Agulnik *et al.*, 1998) and in intron 7A of the *DAZ2* gene (sY586) (Saxena *et al.*, 2000). Because both were also found in the RPCI-11-*DAZ*-BAC sequences, we included them in our SNV marker set as *DAZ*-SNV II and *DAZ*-SNV III respectively. We also included the sY587 marker of Saxena *et al.* as *DAZ*-SNV V for analysis of the *Dra*I restriction site in the *DAZ1* and *DAZ2* genes (Saxena *et al.*, 2000).

Additionally, we established three novel *DAZ*-SNVs (SNV I, SNV IV, SNV VI). SNV I was identified 90 nucleotides upstream of the putative transcription initiation site of the *DAZ* genes with a polymorphic *Fsp*I restriction site only present in *DAZ4*. SNV IV was found 1159 nucleotides downstream of *DAZ* exon 7*H* with a polymorphic *Alu*I restriction site absent in *DAZ2*. SNV VI was found 480 nucleotides downstream of *DAZ* exon 10, where due to a 4 bp deletion (CCTG) at the 3′end of *DAZ4*, an *Afl*III site was created. An extension of this deletion to eight nucleotides (*TCCTGTAT*) was identified at the 3′ end of the *DAZ3* gene and used for the development of a *DAZ3* specific STS (Y-*DAZ3*). A second STS (*DAZ*-RRM3) was established in the repeated intron 1–intron 6 structure specific for *DAZ1* and *DAZ4*. This marker is in the same intronic region as sY152 (Figure 1B) declared earlier to be a marker for the *DAZ1* and *DAZ4* genes (Saxena *et al.*, 2000). However, we found the same sY152 primer pair sequences in the RPCI-11 *DAZ2* and *DAZ3* gene sequences (3′ to exon 7*A* and 7*B* respectively) with only three internal mismatches in the forward primer. Therefore, we often obtained false positive amplification products in our PCR experiments and developed the *DAZ*-RRM3 STS as a more stringent alternative for the identification of the *DAZ1* and *DAZ4* gene copies using our PCR multiplex conditions. All novel *DAZ*-SNVs and STSs including the optimal PCR conditions were deposited in GenBank (accession numbers in Table I).

On the Y chromosome of men with normal fertility, we expect the presence of four *DAZ* gene copies to be indicated by the presence of A and B alleles for each of the six *DAZ*-SNVs as well as the presence of the STSs *DAZ*-RRM3 and Y-*DAZ3*, as shown in Figure 1B. Absence of all *DAZ* markers would indicate the absence of the complete *DAZ* gene family on the Y chromosome. The deletion of a specific *DAZ* gene copy would then be recognized by the absence of specific SNV A or B alleles associated with a particular *DAZ* gene copy and the deletion pattern of the *DAZ*-RRM3 and Y-*DAZ3* STSs (Figure 1B), i.e. by a specific *DAZ* deletion haplotype. In this way, we propose 13 specific *DAZ* deletion haplotypes (*DAZ*-del-haps), each one indicative for the deletion of a specific set of *DAZ* gene copies (Table II).

DAZ gene deletions in men with severe oligozoospermia and in men with proven fertility

We analysed the presence of each of the 13 *DAZ*-del-haps in the genomic DNA samples of 63 men with severe oligozoospermia (OZ samples) and of 107 men with proven fertility (FM samples). All OZ samples had a normal karyotype, 46XY, and Yq11 microdeletions analysed with the STS marker set established previously for the detection of complete AZFa, AZFb and AZFc deletions (Vogt *et al.*, 1996) were not detectable. Four *DAZ*-del-haps (#2, 4, 5, 9) could be identified. Surprisingly, most of them were found in both groups (i.e. fertile and infertile men) with similar frequencies (Tables III and IV). *DAZ*-del-hap 2, indicative for a deletion of *DAZ2*, was found in 26 FM samples (Table III) and in eight OZ samples (Table IV) and is associated with the absence of the B allele for SNV VI, which would suggest an additional deletion in the distal part of *DAZ4* including the SNV VI *Afl*III

Table II. Thirteen *DAZ*-deletion haplotypes are specified by eight *DAZ* SNV/STS markers

DAZ gene deletions	SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	*DAZ*-RRM3	Y-*DAZ3*	*DAZ* del. haplotype
DAZ1	A+B	**B**	A+B	A+B	A+B	A+B	+	+	1
DAZ2	A+B	A+B	**B**	**B**	A+B	A+B	+	+	2
DAZ3	A+B	A+B	A+B	A+B	A+B	A+B	+	–	3
DAZ4	**A**	A+B	A+B	A+B	A+B	**A**	+	+	4
DAZ1+DAZ2	A+B	**B**	**B**	**B**	**A**	A+B	+	+	5
DAZ1+DAZ3	A+B	**B**	A+B	A+B	A+B	A+B	+	–	6
DAZ1+DAZ4	**A**	**B**	A+B	A+B	A+B	**A**	–	+	7
DAZ2+DAZ3	A+B	A+B	**B**	**B**	A+B	A+B	+	–	8
DAZ2+DAZ4	**A**	A+B	**B**	**B**	A+B	**A**	+	+	9
DAZ3+DAZ4	**A**	A+B	A+B	A+B	**B**	**A**	+	–	10
DAZ1+DAZ2+DAZ3	**B**	**B**	**B**	**B**	**A**	**B**	+	–	11
DAZ1+DAZ2+DAZ4	**A**	**B**	**B**	**B**	**A**	**A**	–	+	12
DAZ2+DAZ3+DAZ4	**A**	**A**	**B**	**B**	**B**	**A**	+	–	13

Letters marked in bold indicate the presence of deletions.

Table III. DAZ-del-hap analyses in DNA samples of 107 men with proven fertility

SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	DAZ-RRM3	Y-DAZ3	FM samples	N
A+B	A+B	A+B	A+B	A+B	A+B	+	+	1,4,5,6,7,10,17,20,23,28,31,33,34,38,41,4248,49,50,54, 55,56,57,58,*4,5,6,9,11,13,14,15,16,17,18,19,28,29,30,31, 32,33,34,37,42,43,44,45,46,47*[1]	50
DAZ-del-hap 2 and derivatives									
A+B	A+B	**B**	**B**	A+B	**A**	+	+	9,12,18,19,29,36,46,47,53,60,*1,2,7,10,12,20,22,24,26,35, 36,38,40,48,50*	25
							–	52	1
A+B	A+B	**B**	A+B	A+B	A+B	+	+	8	1
A+B	A+B	**B**	A+B	A+B	**A**	+	+	22,32,59,*27*	4
A+B	A+B	A+B	**B**	A+B	A+B	+	+	51	1
A+B	A+B	A+B	**B**	A+B	**A**	+	+	3,24,44,61,*23*	5
	–							43, 45, *51*	3
DAZ-del-hap 4 and derivatives									
A	A+B	A+B	A+B	A+B	**A**	+	–	39,40	2
A	A+B	A+B	A+B	A+B	A+B	+	+	11,13,25,27,37,*21*	6
A+B	A+B	A+B	A+B	A+B	**A**	+	+	2,15,26	3
DAZ-del-hap 9 and derivatives									
A	A+B	**B**	**B**	A+B	**A**	+	+	16,*8*	2
							–	*25*	1
A	A+B	**B**	A+B	A+B	**A**	+	+	30	1
A	A+B	A+B	**B**	A+B	**A**	+	+	14	1
							–	*3*	1

[1]Italic numbers indicate the 51 Danish controls who had confirmation of a normal sperm parameter.
Letters marked in bold indicate the presence of deletions.

restriction site (Figure 1B). Derivatives of *DAZ*-del hap 2, i.e. absence of only the A allele of SNV III, or only the A allele of SNV IV (with or without absence of the B allele of SNV VI and absence of the Y-*DAZ3* STS marker), were also identified in both groups of men (Tables III and IV). A similar pattern of SNV and STS deletion events was also observed in both sample groups for *DAZ*-del hap 4 (indicative for deletion of *DAZ4*) and *DAZ*-del-hap 9 (indicative for deletion of *DAZ2* and *DAZ4*) (Tables III and IV). It strongly suggests that these *DAZ* deletions are not associated with the occurrence of the oligozoospermic phenotype in the OZ individuals, but are most likely familial variants, inherited from father to son. We were able to confirm this by also finding *DAZ*-del-hap 2 in the fathers of FM 46 and FM 53 and *DAZ*-del-hap 4 in the father of FM 39.

Contrasting with these results, we found a deletion of the *DAZ1* gene linked to a deletion of the *DAZ2* gene (i.e. *DAZ*-del-hap 5; Table II) in only five DNA samples of the oligozoospermic men (OZ 5, 20, 35, 65 and 73; Table IV; Figure 2). Two of them (OZ 20, 35) additionally displayed absence of the A allele for SNV I and of the B allele for SNV VI (Figure 2). This would indicate additionally partial deletions in the 5′ region of the *DAZ3* gene and in the 3′ region of the *DAZ4* gene. In OZ 5, we also found a deletion of the *DAZ*-RRM3 marker (Table IV), which would suggest that the internal part of the repetitive *DAZ4* RRM region (location of

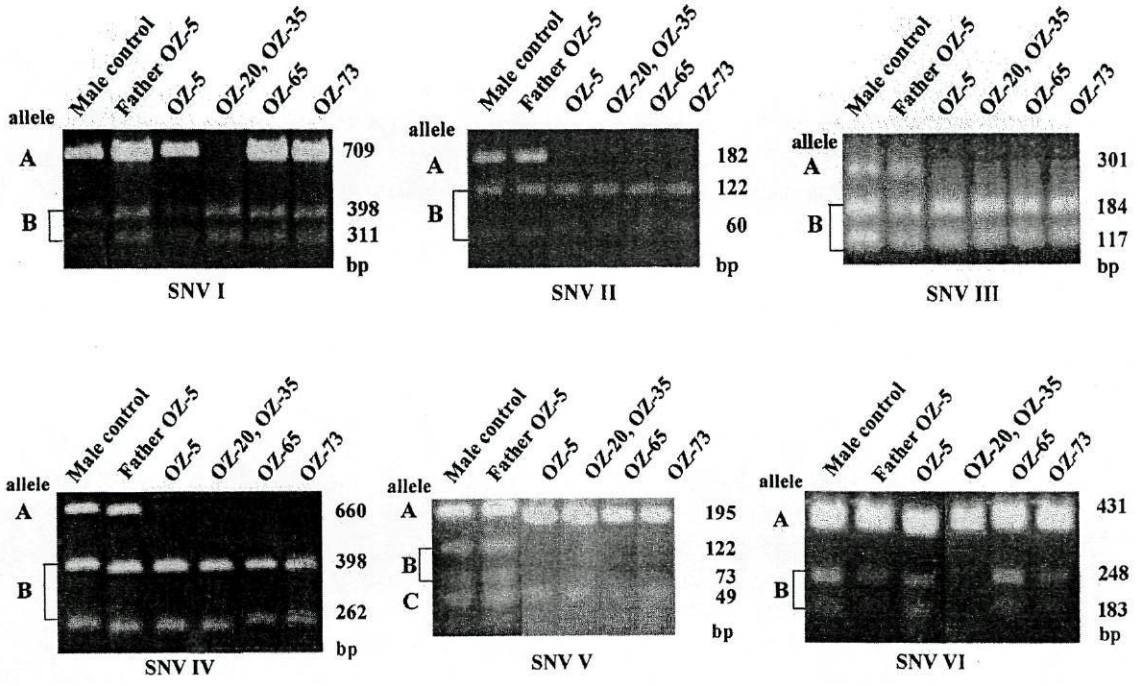


Figure 2. *DAZ*-SNV analyses of genomic DNA samples with *DAZ*-del-hap 5 (OZ 5, OZ -20, OZ -35, OZ -65 and OZ -73). The A allele was defined as not restricted and the B allele as restricted (two fragments). The C fragment in SNV V (49 bp) was constant in all samples analysed. The fragment lengths for each SNV are given in bp on the right side of the pictures. The father of OZ 5 has both alleles for all SNVs tested, as did the male control sample. For further details see text.

Table IV. DAZ-del-hap analyses in genomic DNA samples of 63 oligozoospermic men

SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	DAZ-RRM3	Y-DAZ3	OZ samples	N
A+B	A+B	A+B	A+B	A+B	A+B	+	+	2,3,4,9,11,14,18,22,26,29,30,32,33,34,36,38,40, 41,43,45,46,48,49,50,51,55,56,59,61,68,69,71	32
DAZ-del-hap 2 and derivatives									
A+B	A+B	**B**	**B**	A+B	**A**	+	+	1,15,23,42,52,57,63,77	8
A+B	A+B	**B**	A+B	A+B	**A**	+	+	17,21	2
A+B	A+B	A+B	**B**	A+B	A+B	+	+	66	1
A+B	A+B	A+B	**B**	A+B	**A**	+	+	13,44,47,53	4
							–	24,27,54,62,64	5
DAZ-del-hap 4 and derivatives									
A	A+B	A+B	A+B	A+B	A+B	+	+	39,67	2
A+B	A+B	A+B	A+B	A+B	**A**	+	+	25,37,58	3
DAZ-del-hap 9 and derivatives									
A	A+B	**B**	A+B	A+B	A	+	+	60	1
DAZ-del-hap 5 and derivatives									
A+B	**B**	**B**	**B**	**A**	A+B	+	+	65,73	2
A+B	**B**	**B**	**B**	**A**	A+B	–	+	5	1
B	**B**	**B**	**B**	**A**	**A**	+	+	20,35	2

Letters marked in bold indicate the presence of deletions.

the *DAZ*-RRM3 marker; Figure 1B) is absent in this patient in addition to the deletion of the *DAZ1* and *DAZ2* genes (Table IV). DAZ-del-hap 5 occurred in five of our 63 OZ samples (8%). It could be shown as a 'de-novo' mutation event in the family of OZ 5 as the father of OZ 5 displayed the same complete DAZ-SNV pattern as our fertile male control (Figure 2). Our request for a genomic DNA sample from the fathers of OZ 20, 35, 65 and 73 was unsuccessful because they were deceased or did not want to collaborate. The alternative use of DNA samples from fertile brothers was not possible either, because no brothers were present in these families.

Specific DYS1-EcoRV and -TaqI DNA blot patterns in men with DAZ gene deletions

We found that the deletion of single *DAZ* gene copies could also be identified by DYS1-DNA blot experiments after

restriction digest of genomic DNA samples with *Eco*RV and *Taq*I. A detailed *DAZ* gene copy specific restriction analysis on the sequences of the RPCI-11-*DAZ* BAC contigs revealed that *Eco*RV restriction fragments specific for *DAZ1* and *DAZ4* and *Taq*I restriction fragments specific for *DAZ2* and *DAZ3* were present in the repetitive *DAZ* gene regions (Figure 1B). They are part of the well-known DYS1 locus (Ngo *et al.*, 1986) because they cross-hybridize with the DYS1 cognate fragment 49f (subclone from cosmid 49) (Bishop *et al.*, 1984). Our sequence analysis identified 49f as an *Eco*RI fragment of the *DAZ1* gene, starting 5′ in intron 3 of the third *DAZ1* RRM domain (BAC140H23 sequence position: 34027) and ending 3′ in *DAZ1* intron 7B (BAC140H23 sequence position: 31181; Figure 1B; GenBank accession no. AF414183).

From these findings, we expect a deletion of the DYS1-*Eco*RV fragments 10845 and 10843 bp to be associated with a deletion of the *DAZ1* gene, and a deletion of the 7327 bp DYS1-*Eco*RV fragments to be associated with a deletion of the *DAZ4* gene (Figure 1B). Similarly, a deletion of the DYS1-*Taq*I fragment of 3112 bp should be associated with a deletion of the *DAZ2* gene, and a deletion of the DYS1-*Taq*I fragment of 19612 bp should be associated with a deletion of the *DAZ3* gene. The *Mlu*I restriction patterns identified as specific for each *DAZ* gene copy by Saxena *et al.* (Saxena *et al.*, 2000) could not be used for genomic *DAZ* gene deletion analysis, because the methylation of most *Mlu*I sites in the genomic *DAZ* DNA inhibited restriction (S.Fernandes, unpublished results).

Examples with relevant DNA blot experiments using genomic DNA samples of patients with *DAZ*-del-hap 5 are shown in Figure 3A. A deletion of the *DAZ1* and *DAZ2* genes was confirmed in all *DAZ*-del-hap 5 samples by absence of the *DAZ1*-specific 10.8 kb-*Eco*RV and the *DAZ2*-specific 3.1 kb-*Taq*I fragments of the DYS1 locus. Interestingly, both OZ 5 and his father displayed additionally a familial deletion of *DAZ4*, detected through the absence of the *DAZ4*-specific 7.3 kb *Eco*RV fragment in both DNA samples (Figure 3A). This familial *DAZ4* deletion includes most likely the entire repetitive *DAZ4* gene structure, as the signal intensities of the 5 and 2.4 kb *Eco*RV fragments are weak in comparison with the normal male control (Figure 3A). These two DYS1-*Eco*RV fragments are not gene copy-specific but span the repetitive *DAZ* exon 7 structure in each *DAZ* gene. The presence of the *DAZ4*-specific B alleles of SNV I and SNV VI in OZ 5 and his father (Figure 2), marking the 5′ and 3′ end of the *DAZ4* gene, suggests that the *DAZ4* deletion in this family is partial.

We also performed a DYS1 blot analysis for some DNA samples of men with *DAZ*-del-hap 2 (FM 18, OZ 1 and 15; DYS1-*Taq*I) and *DAZ*-del-hap 9 (FM 16, OZ -60; DYS1-*Taq*I and *Eco*RV). We confirmed a deletion of *DAZ2* through absence of the *DAZ2*-specific 3.1 kb *Taq*I fragment in all of them (Figure 3B), but we could not identify a deletion of the *DAZ4*-specific 7.3 *Eco*RV fragment in FM 16 and OZ 60.

Discussion

Thanks to the Human Genome Sequencing project, it was possible to cover the genomic Y sequence of the AZFc region with a contig of 44 BAC sequences and to estimate its molecular extension to be 4.94 Mb (Figure 1A). This starts at STS sY142 and ends at STS sY167 in distal Yq11.23. sY167 was mapped not >100 kb proximal to the DYZ1 border of the Yq12 heterochromatin, found in BAC428D10 (Figure 1A). A similar BAC contig of the AZFc region was recently published by Kuroda-Kawaguchi *et al.* (Kuroda-Kawaguchi *et al.*, 2001). Their data completely confirm the correct assembly of our AZFc-BAC contig in Figure 1. They extended this analysis by additional iterative alignment processes and revealed a unique repetitive and palindromic AZFc sequence structure along 93% of its length.

AZFc deletions are the most common known genetic cause of human male infertility and are found in different human populations with a rate of ~10^{-4} (Vogt, 1998). In spite of this, it is not yet known whether the Y gene families mapped to this chromosomal region, and which are expressed exclusively in testicular tissue (Lahn and Page, 1997; Vogt, 1998; Kuroda-Kawaguchi *et al.*, 2001) are indeed essential for human spermatogenesis. Gene-specific deletions or point mutations have not been reported in any of the proposed AZFc candidate genes. Therefore it has been proposed that the observed AZFc testicular pathology might only be caused by deletion of the entire AZFc gene content of the human Y chromosome (Vogt, 1998).

In this paper, we present strong experimental evidence that this might not be the case since we found that, in five of 63 OZ individuals, deletions of the *DAZ1/DAZ2* gene doublet are most likely associated with an oligozoospermic phenotype similar to that observed after deletion of the entire AZFc region. Our results are a major step towards the solution of the frequently discussed enigma on the function of the different *DAZ* gene copies in human spermatogenesis.

Variable numbers of DAZ gene copies in the AZFc region of men with proven fertility?

The number and structure of the *DAZ* genes on the human Y chromosome has been assumed to be variable, in a range between three and seven gene copies (Yen *et al.*, 1997; Gläser *et al.*, 1998; Saxena *et al.*, 2000). We did not find more than four *DAZ* gene copies in the RPCI-11 BAC library, nor the presence of truncated *DAZ* pseudogenes. The Y chromosome of the RPCI-11 BAC-library donor should therefore contain four *DAZ* gene copies. They were mapped as *DAZ1/DAZ2* and *DAZ3/DAZ4* doublets in the middle of the 4.94 Mb AZFc BAC contig with a distance of 1.47 Mb between them (Figure 1A). This genomic AZFc structure was first proposed in the AZFc duplication model by Kirsch *et al.* (Kirsch *et al.*, 1996). The presence of four *DAZ* genes in AZFc with the same doublet structure was first found by Saxena *et al.*, although with different repetitive exon 7 domains in *DAZ2*, *DAZ3* and *DAZ4* (Saxena *et al.*, 2000). We also found the four copy gene structure found on the Y chromosome of the RPCI-11 donor in 50 of our 107 fertile control individuals and in 32 of our 63 OZ samples, i.e. in 48% of the samples analysed. That is indicated by the same DAZ-SNV and STS patterns (Tables III and IV). Therefore, we predict that the 4-*DAZ*-gene pattern found in the RPCI-11-*DAZ*-BACs represents the most common *DAZ* gene structure in Caucasian male populations.

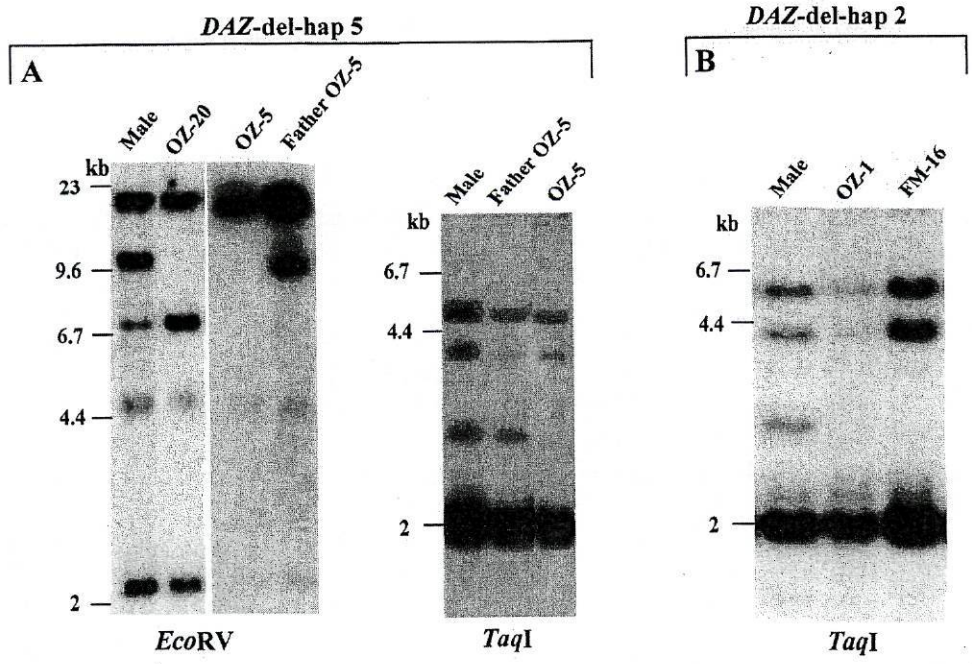


Figure 3. DYS1-*Eco*RV and -*Taq*I DNA blot *DAZ* deletion analyses with the DYS1 cognate probe 49f. (**A**) Example of an *Eco*RV DNA blot from oligozoospermic men identified with *DAZ*-del-hap 5 (OZ -20, OZ 5). *DAZ1* deletion is indicated by absence of the *DAZ1*-specific 10.8 kb *Eco*RV fragment (see also Figure 1B). The same DYS1-*Eco*RV deletion pattern was observed in OZ -35, OZ -65 and OZ -73. A familial *DAZ4* gene deletion in OZ 5 and his father was identified by the additional deletion of the 7.3 kb *Eco*RV fragment in both DNA samples. The DYS1-*Taq*I blot of OZ 5 and his father shows the de-novo deletion of *DAZ2* in OZ 5, indicated by absence of the 3.1 kb *Taq*I fragment only in the OZ 5 DNA sample. The same DYS1-*Taq*I deletion pattern was found for all other OZ samples with *DAZ*-del-hap 5. (**B**) DYS1-*Taq*I blot deletion pattern of OZ 1 (*DAZ*-del-hap 2) and FM 16 (*DAZ*-del-hap 9). A deletion of *DAZ2* is indicated by absence of the 3.1 kb DYS1-*Taq*I fragment. The same DYS1-*Taq*I deletion pattern was found for FM 18 and OZ 15 (*DAZ*-del-hap 2) and OZ 60 (*DAZ*-del-hap 9). In the lanes of OZ 1 and FM 16, different intensities of the autoradiographic signal are due to different DNA amounts. Molecular lengths of the λ-*Hind*III marker that was used are given on the left.

In addition, our *DAZ*-deletion analyses suggest that at least in some individuals with *DAZ*-del-hap 2, two of our two sample groups (FM and OZ), the *DAZ* copy number is reduced to three by deletion of the *DAZ2* gene copy (Figure 3B). The presence of only three *DAZ* gene copies on the Y chromosome of some fertile men was first proposed by Yen *et al.* (Yen *et al.*, 1997). We could not perform a *Taq*I-DYS1 blot experiment with all DNA samples carrying the *DAZ*-del-hap 2 listed in Tables III and IV, because the DNA amount was insufficient and a request for a second genomic DNA sample was not possible; for ethical reasons most fertile male samples had to be collected anonymously, since these individuals had no health problem. We were unable to confirm the additional deletion of *DAZ4* in FM 16 and OZ -60 with *DAZ*-del-hap 9, since the *DAZ4*-specific 7.3 kb *Eco*RV fragment of the DYS1 locus was present in both DNA samples. DNA samples with *DAZ*-del-hap 4, indicative for deletion of only *DAZ4*, were found in the fertile men exclusively (Table III). For the practical reasons given above, DYS1 blot experiments were not possible for these individuals. However, the *DAZ4* deletion in the father of OZ 5 (Figure 3A) and in the father of FM 39 strongly suggests that *DAZ4* deletions are common variants in the normal fertile male population and are not associated with the occurrence of male infertility.

Our *DAZ*-SNV/STS analyses also suggest that the B alleles of SNV I and VI in *DAZ4* are polymorphic. Although they are present in the *DAZ4* gene of the RPCI-11-donor and of 88 individuals in our two sample groups (Tables III and IV), absence of the B allele of SNV VI was found, independently from absence of the B allele of SNV I, in all DNA samples with *DAZ*-del-hap 2 and in three FM -DNA samples with *DAZ*-del-hap 4 (FM 2, 15, 26). Absence of the B allele for SNV I was found, independently from absence of the B allele of SNV VI in six FM and two OZ samples of *DAZ*-del-hap 4 (Tables III and IV). Similar derivatives of *DAZ*-del-hap 2 (i.e. absence of the A allele of only SNV III, or of only SNV IV) were found in FM and OZ samples as well (Tables III and IV). We therefore conclude that in DNA samples where the deletion of *DAZ2* or *DAZ4* could not be confirmed by appropriate DYS1 blot experiments, most probably familial gene conversion events occurred between the highly homologous *DAZ* gene copies (<99%) in the *DAZ* locus of the Y chromosome in these families. Gene conversion events, especially within the *DAZ2* and *DAZ4* gene structures, would easily explain the complex pattern of derivatives which we observed for *DAZ*-del-hap 2, 4 and 9 in both groups of men. Similar conclusions were drawn recently for other human gene loci, based on an unexpected linkage disequilibrium between closely spaced SNP markers (Ardlie *et al.*, 2001). Obviously, gene conversion events prevent the identification of a specific *DAZ* copy deletion by only SNV/STS analyses and the additional analysis by DYS1 blot experiments is required.

We do not think that our findings would be improved by additional fibre–fluorescence in-situ hybridization (FISH)

experiments, because FISH probes including the repetitive *DAZ* gene regions can also result in an over-estimation of the number of *DAZ* gene copies (Saxena *et al.*, 2000) and FISH probes specific for each *DAZ* gene copy cannot be created because of their high sequence homology (<99%). Moreover, irrespective of whether real *DAZ* deletions or simple SNV/STS polymorphisms due to gene conversion events had led to the derivatives of the *DAZ*-del-hap 2, 4, 9 classification in a portion of our FM and OZ samples, both molecular events would not interfere with the men's fertility, as they were found in both sample groups.

Only DAZ1/DAZ2 deletions were associated with severe oligozoospermia

A possible link of a specific *DAZ*-deletion haplotype specific to the occurrence of the patients' oligozoospermic phenotype was found only for the OZ samples with *DAZ*-del-hap 5, pointing to a deletion of the *DAZ1* and *DAZ2* gene copies in the Y chromosome of these men. These were confirmed by the DYS1-DNA blot analysis (Figure 3A). *DAZ*-del-hap 5 was not found in the DNA samples of 107 fertile men nor in the father of OZ 5 (Figure 2, 3A). We therefore assume that not only in OZ 5, but in all OZ samples with *DAZ*-del-hap 5, the *DAZ1/DAZ2* deletions are most likely 'de-novo' mutation events, i.e. restricted to the *DAZ* locus of the patient. A spermatogenic influence of the *DAZ4* deletion in the OZ 5 family can be excluded because the father of OZ 5 was reported to have normal fertility (J.Goncalves, personal communication). No other AZFc STS deletions nor the deletion of any other AZFc candidate genes were identified in the Y chromosome of the *DAZ*-del-hap 5 patients, although small deletions or point mutations in the other AZFc candidate genes cannot yet be excluded.

Our extensive analyses of *DAZ* gene copy specific restriction fragments in the RPCI-11 *DAZ*-BAC contigs (Figure 1B) were only successful for the repetitive DYS1-*DAZ* gene regions. The exact molecular extensions of the indicated *DAZ1/DAZ2* gene deletions could therefore not be estimated. However, absence of the *DAZ1*-specific DYS1-*Eco*RV fragments in all *DAZ*-del-hap 5 patients includes a deletion of at least two of the *DAZ1* RNA binding domains (RRM) in exons 2–5 (Figure 1B). A distortion of the putative DAZ1 germ line-specific RNA binding function is therefore expected. The most probable explanation for the occurrence of the oligozoospermic phenotype in OZ 5, 20, 35, 65 and 73 is therefore the deletion of the entire *DAZ1/DAZ2* gene doublet in all of them.

Putative molecular mechanisms for the DAZ gene deletions in the AZFc region

Our comparison of the *DAZ* gene structures of the RPCI-11 donor with that of the CTA/CTB donors described by Saxena *et al.* reflects a dynamic structure, especially of the *DAZ2*, *DAZ3* and *DAZ4* gene copies, but not of the *DAZ1* gene copy (Saxena *et al.*, 2000). We found that the *DAZ* gene doublets *DAZ1/DAZ2* and *DAZ3/DAZ4* are located in two palindromic sequence domains, separated by a distance of 1.47 Mb (Figure 1A). The observed dynamic rearrangements which distinguished the *DAZ2*, *DAZ3* and *DAZ4* gene copies of the CTA/CTB- and the RPCI-11 donors, as well as the proposed gene conversion events discussed above, might therefore be caused by the structural-specific organization in the genomic region of both *DAZ* gene doublets and/or of the entire AZFc sequence, now shown to be composed of massive palindromes in 93% of its length (Kuroda-Kawaguchi *et al.*, 2001). Palindromes and inverted repeat structures in the genome are generally known to induce local instabilities which depend on their lengths and repetitive sequence contents. Local rearrangements increasing their asymmetry could stabilize them (Zhou *et al.*, 2001). In the *DAZ1/DAZ2* and *DAZ3/DAZ4* palindromes, this could explain the amplification of RRM domains in only *DAZ1* and *DAZ4* and the variable structure of the repetitive exon 7 in *DAZ2*, *DAZ3* and *DAZ4*.

Deletions of the *DAZ1* gene alone were found neither in our fertile men nor in our infertile men, but only in combination with the *DAZ2* gene. We therefore assume that molecular mechanisms inducing the observed *DAZ1/DAZ2* gene deletion events are most likely based on the nature of their unique palindromic sequence structure or on the recombination of homologous repetitive sequence blocks flanking the *DAZ1/DAZ2* gene doublet (Chen *et al.*, 1997). Intrachromosomal recombination events of two repetitive 10 kb sequence blocks in proximal Yq11 were recently found to be causative agents for the occurrence of most AZFa deletions (Kamp *et al.*, 2000; Sun *et al.*, 2000). Long homologous repetitive sequence blocks were also found around the *DAZ* genes. Two long repetitive blocks (AZFc-Rep1, Figure 1) in the same polarity and with a surprisingly high sequence homology (<99%), starting in BAC5C5 and BAC160O2 (respectively positions 134890 and 50000) and extending to the 3′ends of the *DAZ1* and *DAZ3* gene respectively, were easily identified by BLAST alignment (Figure 1A). Their molecular lengths were estimated from the AZFc-BAC contig sequence to be 368 (*DAZ1*-AZFc-Rep1) and 320 kb (*DAZ3*-AZFc-Rep1). Different amplifications of the human β-satellite 68 bp repeat unit (Waye *et al.*, 1989) explain their length differences. Intrachromosomal recombination events between these two AZFc-Rep1 blocks would delete not only the *DAZ1* gene, but also the entire *DAZ2* gene, as demonstrated in our five OZ samples with *DAZ*-del-hap 5.

We also found a considerable extension of the palindromic inverted repeats wherein the *DAZ1/DAZ2* and *DAZ3/DAZ4* gene doublets were embedded. These structures would certainly also promote a deletion of the complete gene doublets (Zhou *et al.*, 2001).

The presence of the long repetitive AZFc-Rep1 sequence block with the same polarity upstream of the *DAZ1/DAZ2* and *DAZ3/DAZ4* gene doublets respectively was confirmed by the work of Kuroda-Kawaguchi *et al.* (Kuroda-Kawaguchi *et al.*, 2001). In their paper, the AZFc-Rep1 blocks are the green AZFc amplicons, g1 and g2. A third homologous amplicon, g3, but with an opposite polarity, was found 3′ downstream of the *DAZ3/DAZ4* gene doublet (Kuroda-Kawaguchi *et al.*, 2001; Figure 1 in that paper). Based on this palindromic AZFc sequence structure, we indeed expect that a deletion of the complete *DAZ1/DAZ2* gene doublet can be caused by intrachromosomal recombination events between the AZFc-Rep1 blocks as shown in Figure 1, or, as a second possibility, between the

AZFc amplicons b1 and b3 (Kuroda-Kawaguchi *et al.*, 2001). *DAZ1/DAZ2* deletions caused by b1/b3 recombination events would be indicated by the absence of two AZFc-STSs: DYS227/sY139 (GenBank accession no. G12002) mapped to AZFc-u2 (S.Fernandes, unpublished results) and sY1192 (GenBank accession no. G67166) mapped to AZFc-u3 (Kuroda-Kawaguchi *et al.*, 2001). *DAZ1/DAZ2* deletions caused by AZFc-Rep1 (or g1/g2) recombination events would be indicated by absence of the P1.2 amplicon of the AZFc-yellow-palindrome (Kuroda-Kawaguchi *et al.*, 2001). Therefore, for the analysis of AZFc-Rep1 recombination events we established a novel SNV in the AZFc-P1.1/P1.2 amplicons (AZFc-yellow-palindrome-SNV I; GenBank accession no. G73351). Only AZFc-P1.2 located between g1/g2 has the A allele and AZFc-P1.1 has the B allele (*Dde*I restriction site; S.Fernandes, unpublished results).

We exclude a b1/b3 recombination event as a causative agent for the proposed complete deletion of the DAZ1/DAZ2 gene doublet in the OZ samples with DAZ-del-hap 5 (OZ 5, 20, 35, 65 and 73) because STS sY139 and sY1192 were found to be present in all these DNA samples. However, absence of the A allele for SNV I of the AZFc-yellow-palindrome was noted in OZ 5 and OZ 65, indicating deletion of AZFc-P1.2. This confirms our assumption of a AZFc-Rep1 recombination event as the causative agent for a complete deletion of the *DAZ1/DAZ2* gene doublet in these OZ samples. Interestingly, the three other OZ samples with DAZ-del-hap 5 (OZ 20, 35 and 73) displayed a deletion of the SNV I B allele of the AZFc-yellow-palindrome, suggesting a deletion of the P1.1 amplicon in these samples. Although not yet proven, we speculate that in the Y chromosome of these three OZ families both AZFc-yellow amplicons (P1.1 and P1.2) have an inverted orientation when compared with that in the AZFc sequence of the Y chromosome of the RPCI-11 donor (Kuroda-Kawaguchi *et al.*, 2001). If this holds true, we can suggest that a general reference for the AZFc sequence as proposed by Kuroda-Kawaguchi *et al.* might not exist (Kuroda-Kawaguchi *et al.*, 2001) and that in the OZ 20, 35 and 73 samples, a recombination event between the AZFc-Rep1 blocks has caused a complete deletion of the *DAZ1/DAZ2* gene doublet, as indicated by absence of the B allele of the AZFc-yellow-palindrome SNV I.

Acknowledgements

We are grateful to Hartmut Voss for sequence analysis of the 49f fragment. Drs N.Joergensen, A.G.Andersen and J.Hinrichsen are thanked for their help in collecting and isolating DNA samples from fertile male controls (partially supported by the European Research Commission). J.Riekenberg and A.Barbosa are thanked for their help in collecting the samples from the oligozoospermic men. We also wish to thank Kartrin Christophers-Genzmer and 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 grants to S.F. from the Fundacao para a Ciencia e Tecnologia (FCT) (SFRH/BD/811/2000), to J.G. from the project PRAXIS XXI/PSAU/97/96 of the FCT and to P.H.V. from the Deutsche Forschungsgemeinschaft (DFG: Vo403/10–2).

Note Added in Proof

Recently, a reduced number of DAZ gene copies in subfertile and infertile men has also been reported by de Vries *et al.* (2002, *Fertil. Steril.*, **77**, 68–75), using *DAZ*-SNV deletion and *DAZ*-Fibre-FISH experiments. Surprisingly, the Fibre-FISH experiments did not support their SNV deletion analyses and Fibre-FISH was also not able to distinguish between the deletion of one *DAZ* gene copy and the deletion of a complete *DAZ* gene cluster (*DAZ1/DAZ2* or *DAZ3/DAZ4*, respectively). To confirm partial *DAZ* gene deletions, the use of DYS1 blot experiments instead of Fibre-FISH, as proposed in this paper, is therefore strongly recommended.

References

Agulnik, A.I., Zharkikh, A., Boettger-Tong, H., Bourgeron, T., McElreavey, K. and Bishop, C.E. (1998) Evolution of the *DAZ* gene family suggests that Y-linked *DAZ* plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. *Hum. Mol. Genet.*, **7**, 1371–1377

Ardlie, K., Liu-Cordero, S.N., Eberle, M.A., Daly, M., Barrett, J., Winchester, E., Lander, E.S. and Kruglyak, L. (2001) Lower-than-expected linkage disequilibrium between tightly linked markers in human suggests a role for gene conversion. *Am. J. Hum. Genet.*, **69**, 582–589

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.

Burd, C.G. and Dreyfuss, G. (1996) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

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.

Gläser, B., Yen, P.H. and Schempp, W. (1998) Fibre-fluorescence *in situ* hybridization unravels apparently seven *DAZ* genes or pseudogenes clustered within a Y-chromosome region frequently deleted in azoospermic males. *Chromosome Res.*, **6**, 481–486.

Habermann, B., Mi, H.F., Edelmann, A., Bohring, C., Backert, I.T., Kiesewetter, F., Aumuller, G. and Vogt, P.H. (1998) *DAZ* (Deleted in Azoospermia) genes encode proteins located in human late spermatids and in sperm tails. *Hum. Reprod.*, **13**, 363–369.

Houston, D.W., Zhang, J., Maines, J.Z., Wasserman, S.A. and King, M.L. (1998) A *Xenopus DAZ*-like gene encodes an RNA component of germ plasm and is a functional homologue of Drosophila boule. *Development*, **125**, 171–180.

Kamp, C., Hirschmann, P., Voss, H., Huellen, K. and Vogt, P.H. (2000) Two homologous retroviral sequence blocks in proximal Yq11 cause AZFa mircrodeletions as a result of intrachromosomal recombination events. *Hum. Mol. Genet.*, **9**, 2563–2572.

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.

Kleiman, S.E., Yogev, L., Gamzu, R., Hauser, R., Botchan, A., Paz, G., Lessing, J.B., Yaron, Y. and Yavetz, H. (1999) Three-Generation Evaluation of Y-Chromosome Microdeletion. *J. Androl.*, **20**, 394–398.

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S. and Page, D.C. (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.*, **29**, 279–286.

Lahn, B.T. and Page, D.C. (1997) Functional coherence of the human Y chromosome. *Science*, **278**, 675–680.

Ngo, K.Y., Vergnaud, G., Johnsson, C., Lucotte, G. and Weissenbach, J. (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. *Am. J. Hum. Genet.*, **38**, 407–418.

Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial chromosome libraries. *Genomics*, **51**, 1–8.

Reijo, R.A., DorFM an, D.M., Slee, R., Renshaw, A.A., Loughlin, K.R., Cooke, H. and Page, D.C. (2000) *DAZ* family protein exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. *Biol. Reprod.*, **63**, 1490–1496.

Ruggiu, M., Speed, R., Taggart, M., McKay, S.J., Kilanowski, F., Saunders, P., Dorin, J. and Cooke, H.J. (1997) The mouse *DAZla* gene encodes a cytoplasmic protein essential for gametogenesis. *Nature*, **389**, 73–77.

Saxena, R., Brown, L.G., Hawkins, T., Alagappan, R.K., Skaletsky, H., Reeve, M.P., Reijo, R., Rozen, S., Dinulos, M.B., Disteche, C.M. and Page, D.C.

(1996) The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed repeatedly amplified and pruned. *Nature Genet.*, **14**, 292–298.

Saxena, R., de Vries, J.W., Repping, S., Alagappan, R.K., Skaletsky, H., Brown, L.G., Ma, P., Chen, E., Hoovers, J.M. and Page, D.C. (2000) Four *DAZ* genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics*, **67**, 256–267.

Seboun, E., Barbaux, S., Bourgeron, T., Nishi, S., Agulnik, A., Egashira, M., Nikkawa, N., Bishop, C., Fellous, M., McElreavey, K. *et al.* (1996) A *SPGY* copy homologous to the mouse gene *DAZla* and the Drosophila gene boule is autosomal and expressed only in the human male gonad. *Hum. Mol. Genet.*, **5**, 2005–2011.

Seboun, E., Barbaux, S., Bourgeron, T., Nishi, S., Agulnik, A., Egashira, M., Nikkawa, N., Bishop, C., Fellous, M., McElreavey, K. *et al.* (1997) Gene sequence, localization, and evolutionary conservation of *DAZLA*, a candidate male sterility gene. *Genomics*, **41**, 227–235.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA*, **89**, 8794–8797.

Simoni, M., Bakker, E., Eurlings, M.C., Matthijs, G., Moro, E., Muller, C.R. and Vogt, P.H. (1999) Laboratory guidelines for molecular diagnosis of Y-chromosomal microdeletions. *Int. J. Androl.*, **22**, 292–299.

Slee, R., Grimes, B., Speed, R.M., Taggart, M., Maguire, S.M., Ross, A., McGill, N.I., Saunders, P.T. and Cooke, H.J. (1999) A human *DAZ* transgene confers partial rescue of the mouse *DAZl* null phenotype. *Proc. Natl Acad. Sci. USA*, **96**, 8040–8045.

Sun, C., Skaletsky, H., ROZ en, S., Gromoll, J., Nieschlag, E., Oates, R. and Page, D.C. (2000) Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum. Mol. Genet.*, **9**, 2291–2296

Tilford, C.A., Kuroda-Kawaguchi, T., Skaletsky, H., ROZ en, S., Brown, L.G., Rosenberg, M., McPherson, J.D., Wylie, K., Sekhon, M., Kucaba, T.A. *et al.* (2001) A physical map of the human Y chromosome. *Nature*, **409**, 943–945.

Vogt, P.H. (1998) Human chromosome deletion in Yq11, AZF candidate genes and male infertility: history and update. *Mol. Hum. Reprod.*, **4**, 739–744.

Vogt, P.H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Kohn, 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.

Vogt, P.H., Affara, N., Davey, P., Hammer, M., Jobling, M.A., Lau, Y.F., 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.

Waye, J.S, and Willard, H.F. (1989) Human β satellite DNA: genomic organisation and sequence definition of a class of highly repetitive tandem DNA. *Proc. Natl Acad. Sci. USA*, **86**, 6250–6254.

World Health Organisation (1999) *Laboratory manual for the examination of human semen and sperm-cervical mucus interaction.* Cambridge University Press, Cambridge, UK.

Yen, P.H., Chai, N.N. and Salido, E.C. (1997) The human *DAZ* genes a putative male infertility factor on the Y chromosome, are highly polymorphic in the *DAZ* repeat regions. *Mamm. Genome*, **8**, 756–759.

Zhou, Z.-H., Akgün, E. and Jasin, M. (2001) Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. *Proc. Natl Acad. Sci. USA*, **98**, 8326–8333.

Submitted on November 2, 2001; accepted on January 15, 2002

APMIS 111: 115–27, 2003
Printed in Denmark . All rights reserved

Copyright © APMIS 2003

APMIS
ISSN 0903-4641

Polymorphic DAZ gene family in polymorphic structure of AZFc locus: Artwork or functional for human spermatogenesis?

Review article

PETER H. VOGT and SUSANA FERNANDES

Section Molecular Genetics & Infertility, Dept. Gynecol. Endocrinol. & Reproductive Medicine, University of Heidelberg, D-69115 Heidelberg, Germany

Vogt PH, Fernandes S. Polymorphic DAZ gene family in polymorphic structure of AZFc locus: Artwork or functional for human spermatogenesis? Review article. APMIS 2003;111:115–27.

Human spermatogenesis is regulated by a network of genes located on autosomes and on sex chromosomes, but especially on the Y chromosome. Most results concerning the germ cell function of the Y genes were obtained by genomic breakpoint mapping studies of the Y chromosome of infertile patients. Although this approach has the benefit of focussing on those Y regions that contain most likely the Y genes of functional importance, its major drawback is the fact that fertile control samples were often missing. In fertile men, molecular and cytogenetic analyses of the Y chromosome has revealed highly polymorphic chromatin domains especially in the distal euchromatic part (Yq11.23) and in the heterochromatic part (Yq12) of the long arm. In sterile patients cytogenetic analyses mapped microscopically visible Y deletions and rearrangements in the same polymorphic Y regions. The presence of a Y chromosomal spermatogenesis locus was postulated to be located in Yq11.23 and designated as AZoospermia Factor (AZF). More recently, molecular deletion mapping in Yq11 has revealed a series of microdeletions that could be mapped to one of three different AZF loci: AZFa in proximal Yq11 (Yq11.21), AZFb and AZFc in two non-overlapping Y-regions in distal Yq11 (Yq11.23). This view was supported by the observation that AZFa and AZFb microdeletions were associated with a specific pathology in the patients' testis tissue. Only AZFc deletions were associated with a variable testicular pathology and in rare cases AZFc deletions were even found inherited from father to son. However, AZFc deletions were found with a frequency of 10–20% only in infertile men and most of them were proved to be "de novo", i.e. the AZFc deletion was restricted to the patient's Y chromosome. Based mainly on positional cloning experiments of testis cDNA clones and on the Y chromosomal sequence now published in GenBank, a first blueprint for the putative gene content of the AZFc locus can now be given and the gene location compared to the polymorphic DNA domains. This artwork of repetitive sequence blocks called AZFc amplicons raised the question whether the AZFc chromatin is still part of the heterochromatic domain of the Y long arm well known for its polymorphic extensions or is decondensed and part of the Yq11.23 euchromatin? We discuss also the polymorphic DAZ gene family and disclose putative origins of its molecular heterogeneity in fertile and infertile men recently identified by the analyses of Single Nucleotide Variants (SNVs) in this AZFc gene locus.

Key words: AZF locus; Yq11; DAZ locus; chromatin folding code.

Peter H. Vogt, Section Molecular Genetics & Infertility, Dept. Gynecol. Endocrinol. & Reproductive Medicine, Frauenklinik, University of Heidelberg, Voßstrasse 9, D-69115 Heidelberg, Germany.
e-mail: peter_vogt@med.uni-heidelberg.de

POLYMORPHIC HUMAN Y CHROMOSOME: NO GENES FOR SPERMATOGENESIS?

Human spermatogenesis is regulated by a network of genes located on autosomes and on the sex chromosomes, but especially on the Y chromosome. This has been known in Drosophila for nearly 80 years. In mammals it was revealed only recently, as it was hidden by the sex determination function of the Y chromosome. However, the human Y chromosome seemed to be an exception because its chromatin structure has a highly variable length (1). This length polymorphism could be mapped to the long Y arm after staining the chromosomes with quinacrine (Fig. 1) (2). The length of the fluorescent chromatin block is constant in family pedigrees and therefore useful as a family marker. Despite of these large Y chromosomal rearrangements, no effects on the fertility statues of these men have ever been reported, even when the fluorescent Y marker was completely absent indicating large deletions in Yq. This was unexpected since in other chromosomes deficiencies or duplications of only small segments most often led to severe phenotypic damages. Consequently, the human Y chromosome was first considered to be only functional for sex determination but would have no genes for spermatogenesis.

AZOOSPERMIA FACTOR (AZF) IN Yq11 IS GENETICALLY SPLIT IN 3 SUBINTERVALS (AZFa, AZFb, AZFc)

A straight-forward strategy for the identification of protein encoding Y genes with a putative function in spermatogenesis would be the isolation of Y chromosomal cDNA clones and the study of their expression profiles in human testis tissue. However, in human the isolation of testis tissue for experimental reasons is usually not possible because humans are not experimental species. Therefore, most results supporting a germ cell function of the human Y chromosome were first collected from genomic breakpoint mapping studies of the Y chromosome of sterile patients. Cytogenetic analyses mapped microscopically visible deletions and rearrangements of the Y chromosome found in sterile patients always in Yq11 (3, 4). This Y deletions were characterised by a complete absence of the fluorescent heterochromatin. Monocentric Yq-chromosomes, dicentric double Y chromosomes (iso-Yp), Y with a ring structure (ring-Y), or translocation of the Y to an autosome were most often identified (Fig. 2). A polymorphic Y rearrangement in the patients' families were excluded because the Y chromosomes of their fathers had a normal fluorescent heterochromatin block. These Y deletions could therefore be associated with the occurrence of the men's infertile phenotypes (5). Most of them were azoospermic, meaning that no mature sperm cells were detectable in their semen fluid. Tiepolo and Zuffardi (1976) therefore designated this Y fertility factor as AZoospermia

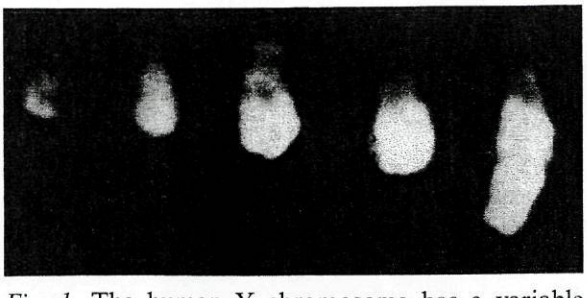


Fig. 1. The human Y chromosome has a variable staining pattern with quinacrine marking the distal long arm with different intensities (2).

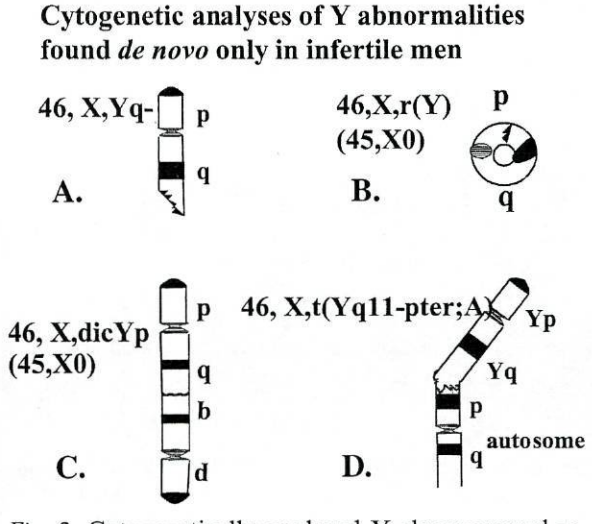


Fig. 2. Cytogenetically analysed Y chromosomal rearrangements and deletions observed in men with idiopathic sterility. Breakpoint-fusion-regions of each of these Y chromosomal rearrangements include part of the euchromatic part of the Y-long arm (Yq11).

Factor (AZF) and mapped it in the distal part of the euchromatic Yq11 region, the G-light band Yq11.23, flanking the heterochromatic chromatin block, since the fluorescent heterochromatin domain was expected to be genetically inert. Nevertheless, because of the observed polymorphic deletion events observed in Yq in fertile men (see Fig. 1), for a long time nobody believed in a functional AZF locus on the polymorphic Y long arm and it was considered to be impossible to distinguish Y chromosomal variabilities which are indeed associated with disruption of an essential Y chromosomal AZF locus.

A novel and more promising approach towards the location and identification of the proposed AZF locus was introduced then by the analyses of molecular microdeletions in the Y chromosomes of sterile men with a normal karyotype (46,XY). In the first large sample study using this molecular approach, deletion mapping in Yq11 revealed a series of microdeletions which all of them could be mapped to one of 3 different genomic Y regions (6). Since they were only observed in men with azoospermia or severe oligozoospermia and were not present in the patients' fathers, it was proposed that not only one but at least 3 AZF loci are present in Yq11: AZFa in proximal Yq11 (Yq11.21), AZFb and AZFc in two non-overlapping Y-regions in distal Yq11 (Yq11.23) (Fig. 3). This view was supported by the observation that AZFa and AZFb microdeletions were associated with the occurrence of a specific pathology in the patients' testis tissue. Complete AZFa deletions were always associated with a complete absence of germ cells also called Sertoli-cell-only (SCO) syndrome. Complete absence of AZFb was always associated with a meiotic maturation arrest, i.e. spermatogonia and spermatocytes were present in the patients' testis tubules in normal amounts but postmeiotic germ cells were completely absent. Only AZFc deletions in distal Yq11 were associated with a variable testicular pathology and in rare cases AZFc deletions were found to be inherited from father to son. Since AZFc was located near the polymorphic heterochromatic Y chromatin block, it was first argued that the patient's AZFc deletion was indeed not associated with the occurrence of infertility in this patient but would be one of the polymorphic deletion events encompassing different lengths of the flanking heterochromatic Y-DNA. However, AZFc deletions are now known as the most common genetic lesion found in men with azoospermia or severe oligozoospermia in a frequency of 10–20%. Moreover, most of them were proven to be "de novo", i.e. most AZFc deletions were restricted to the patient's Y chromosome (7–9).

The fact that each Y microdeletion was part of the AZF locus functional in human spermatogenesis became obvious after the molecular isolation and structural characterisation of its gene content. A first rough estimation of the deleted DNA length in the 3 AZF loci ranged between 1–3 million nucleotides (6), thus large enough to accommodate more than one spermatogenesis gene in each of them. And this is exactly the case. Based on the knowledge of the Y chromosomal sequence, now published in GenBank (10) and extensive testes cDNA screening programs (11–13), a first blueprint for the putative gene content of each AZF locus can be given (Fig. 3): AZFa contains 2 Y genes (*USP9Y* and *DBY*), AZFb 7 Y genes (*CDY2, EIF1AY, PRY, RBMY1, SMCY, TTY5, TTY6*), and AZFc 7 Y genes (*BPY2, CDY1, CSPG4LY, DAZ, GOLGA2LY, TTY3, TTY4*). First the RBMY gene family in AZFb was designated as the most significant AZF gene (14), secondly than the *DAZ* gene family became *en vogue* although first identified only as a single copy gene in AZFc (15). However, as it is now obvious that multiple genes expressed in testis tissue were present in each AZF locus, the designation of the *RBMY* or *DAZ* genes as the only AZF genes can not be preserved anymore and each Y gene of the AZF loci should be considered as a similar strong candidate for the Y-AZF function.

Only the identification of gene-specific mutations found in men with a distinct testicular pathology are suited to reveal the functional state of each of these AZF-Y genes in human spermatogenesis. In AZFa, a 4bp deletion in the splice-donor site of exon 8 of *USP9Y* was identified as "de novo" mutation in one patient with spermatid arrest (16). This suggests a functional contribution of *USP9Y* to human spermatogenesis, but as the AZFa pathology, a complete SCO syndrome, did not occur in this patient, the *DBY* gene or other genetic factors in AZFa

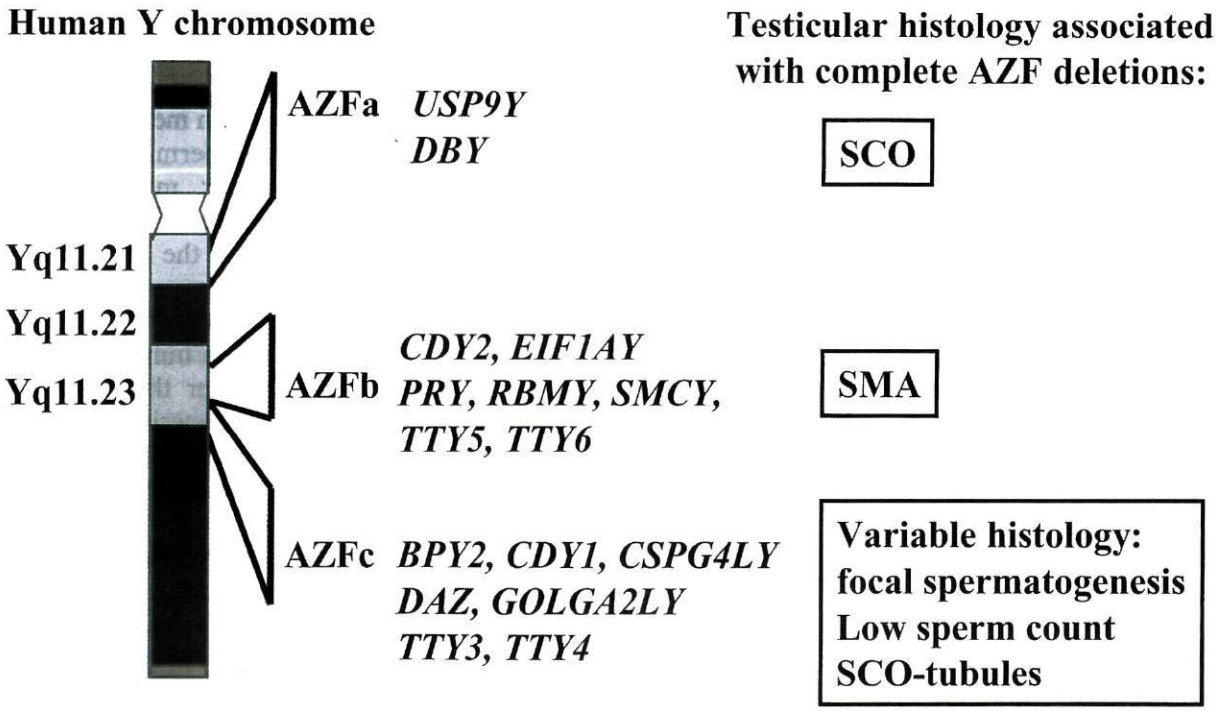


Fig. 3. Three different microdeletions in Yq11 found "de novo" in men with idiopathic infertility split the AZF locus in AZFa associated with the occurrence of a complete Sertoli-cell-only-syndrome (SCO), in AZFb associated with the occurrence of a complete spermatocyte arrest (SMA) and in AZFc associated with the occurrence of a variable testicular pathology. In each AZF microdeletion a series of genes were identified expressed in human testis tissue. For their nomenclature and further details see Lahn & Page 1997; Vogt et al. 1997; Kawaguchi-Kuroda et al. 2001.

must also contribute to the AZFa function. In AZFb, gene-specific mutations were found in the *RBMY* gene locus in infertile men but also in fertile men, revealing a polymorphic nature of the copy number of this gene family (17). In AZFc, gene specific mutations are difficult or even impossible to map because all AZFc genes have multiple copies with a 99.9% sequence homology (see below).

In summary, it is now well established that the human Y chromosome is essential for spermatogenesis and it seems to have evolved a friendly harbour (i.e. the AZF locus) especially for genes expressed in testis tissue. It is most likely that evolutionary benefits have selected this Y chromosomal location for the testis expressed genes probably because only the Y chromosome is male specific. Now it lost most of its ancient sequence domains homologous to the X chromosome and therefore the Y genes are protected from their lost by the decreasing rate of X-Y recombinations along most of the length of Yq11. Since the mouse Y chromosome only contains homologues of the AZFa Y genes (*Usp9y* and *Dby*) and AZFb-Y genes (*Smcy, Eifay1, Rbm*), the AZFc locus can be also considered as an evolutionarily very young Y spermatogenesis locus. All the ancestor genes of its gene families were transposed from different autosomes to the Y chromosome (13) probably first during evolution of the primates (18, 19).

POLYMORPHIC AZFc STRUCTURE IN HUMAN Y CHROMOSOME

Polymorphic lengths of the euchromatic part of the Y long arm (Yq11) were first reported 30 years ago (20) and molecular Y DNA analyses have then mapped 2 DNA loci with polymorphic restriction fragments in AZFc (DYS1; DYS7). The genomic DNA locus DYS7 is recognised by the cognate probe 50f2, which cross hybridises to 5 genomic Y-DNA fragments (A–E), after *Eco*RI restriction (21). 50f2/A,B was mapped to proximal Yp, 50f2/C to proximal AZFc, 50f2/D to the Y centromer and 50f2/E to AZFb (6). After analyses of 859 men from

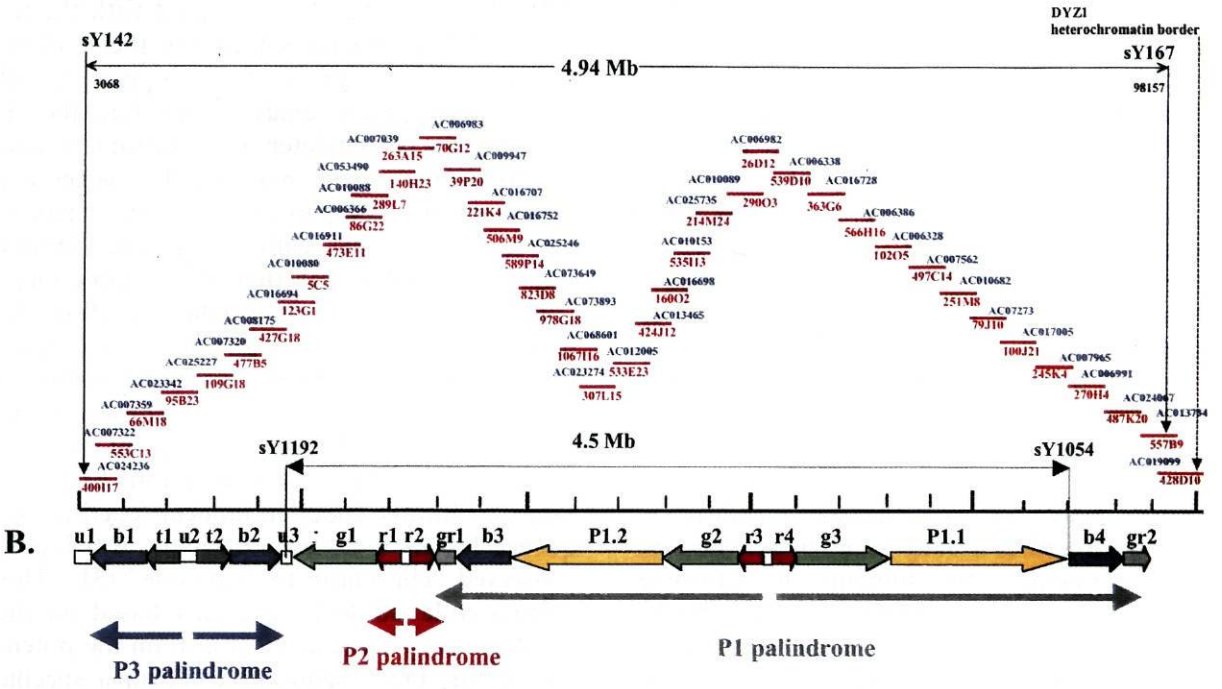


Fig. 4. A. Schematic view of the AZFc RPCI-11 BAC contig between the STS loci sY142 and sY167 (i.e. 4.94 Mb). Their continuous overlap was guaranteed by GenBank data and own BLAST2 alignments of the proposed overlapping end fragments. BAC428D10 including the DYZ1 border of the Yq12 heterochromatin overlapped to the distal AZFc border sequence in BAC557B9 (shown at the right). The GenBank accession number of each BAC sequence (blue colour) is given above the corresponding BAC clone (red colour). B. Schematic view on the repetitive amplicon structure of the AZFc BAC contig according to the colour code of Kawaguchi-Kuroda et al. (2001). With aid of a series of novel STSs (see text) the authors were able to narrow the molecular extension of the AZFc locus to 4.5 Mb. SY1192 and sY1054 were always absent in men with AZFc deletions. The molecular lengths of the 3 palindromes (P1, P2, P3) identified in the repetitive Y amplicon structure are 3Mb for P1, 270 kb for P2 and 990 kb for P3. For further details see text.

46 different populations, absence of the 50f2/C fragment was found in 55 of them and duplication was found in 8 men (22). Obviously, the genomic *Eco*RI fragment in AZFc designated as 50f2/C is part of a polymorphic deletion-insertion DNA locus in AZFc. With knowledge of the AZFc sequence (see Fig. 4) 50f2/C could now be mapped to proximal AZFc since some 50f2/C deletions were found to be associated with deletions of neighboured *RBMY* gene copies but not with *DAZ* gene copies. However, most polymorphic 50f2/C deletions were reported to be smaller not including other DNA loci (22). The frequency of 50f2/C deletions is variable in the 12 human populations analysed and was found with the highest frequency in Finns: 11 of 20 men had this deletion (22). This suggests the occurrence of multiple founder deletions in the proximal AZFc locus including the 50f2/C *Eco*RI fragment.

Interestingly, the first interstitial microdeletion in Yq11 supposed to be associated with the occurrence of germ cell aplasia in the patient's testis tubule was marked by a deletion of the polymorphic genomic 50f2/C fragment (23). Obviously, at that time the authors were not aware of the 50f2/C deletion polymorphism, although they would have easily detected it, if their study would have included an analysis of the Y chromosome of the patient's father.

The molecular structure of the AZFc locus is now established by a contig of 44 BAC clones in distal Yq11 with all the sequences published in GenBank (Fig. 4A). Between the STSs (Sequence Tagged Sites), sY142 and sY167, always present in men with AZFc deletions, the mol-

ecular extension was roughly estimated in 4.94 Mb (24). This was not only confirmed but also improved by similar deletion analyses with a series of novel STSs (sY1054, sY1192, sY1197, sY1125), which further narrowed the extension of the AZFc locus to 4.5 Mb (Fig. 4B). SY1192 and sY1054 were always absent in men with AZFc deletions (13). The molecular extensions of AZFc deletions seem therefore to be rather uniform. With the aid of sophisticated BLAST alignment tools Timoko Kuroda-Kawaguchi was able to show that the complete AZFc sequence consists of a series of homologous large repetitive sequence blocks, which she designated as "amplicons" and which are mainly organised in two palindromes (Fig. 4B). Interestingly, this artwork of repetitive sequence blocks is comparable with the locus-specific repetitive sequence structures usually found in pericentromeric heterochromatin domains of autosomes (25). Since the first DYZ1 block of the Yq heterochromatin block has been mapped immediately adjacent to the distal border of AZFc (24), the question is now raised whether at least part of the chromatin structure of the AZFc locus is still part of the Yq heterochromatin block and not yet part of the decondensed euchromatin domain proposed earlier and marked by the G-light band called Yq11.23 (4).

Several earlier cytogenetic studies have demonstrated that the heterochromatic chromatin domain on the distal long Y arm is not homogeneous, but that it consists of a complex succession of distinct subbands with a narrow gap in the middle of the fluorescent heterochromatin block (26). In C-banded prometaphases the Y chromosome is less condensed and up to five small chromatin domains could be distinguished (27). By restriction with the endonuclease *Taq*I, the proximal part of the Y-heterochromatin could be differentiated from the distal part also indicating a different chromosome structure in the proximal Yq heterochromatin (28) and after incorporation of 5–bromodeoxyuridine the same chromosome region stains differentially with quinacrine plus Giemsa (FPG) (29). Obviously that the heterochromatic chromosome block of the human Y is not uniform and the possibility exists that the proximal heterochromatic chromatin domain distinguished by the earlier cytogenetic Y experiments at least overlaps with the chromatin domain of the repetitive amplicon structure in the AZFc locus. The DNA structure of the AZFc locus can be also compared with the repetitive DNA structures found in the Y chromosomal fertility genes of Drosophila. In all Drosophila species analysed so far, the Y chromosome is completely heterochromatic and the structures of the male fertility genes are composed of a complex organisation of repetitive sequence blocks with a Y-specific "library code" comparable with that of the AZFc locus (30). This chromosome-specific so-called "library code" is a general and conserved structural feature of tandem repetitive sequence blocks and can be found in heterochromatic domains of Drosophila, mouse and plant chromosomes as well (31). It was therefore proposed that the "library code" in tandem repetitive sequence blocks is an essential part of a highly conserved "chromatin folding code" (31). This folding code would be not only based on the DNA primary sequence, but also on the potential of this DNA sequence to form a specific helical secondary structure and on its potential to bind specific proteins which are needed to stabilise a locus-specific chromatin folding structure (32).

We know now that complete AZFc deletions are caused after intrachromosomal recombinations between the b2/b4 amplicons located in a distance of about 4.5 Mb in distal Yq11 (13). However, more large repetitive sequence blocks with the same polarity exist in AZFc (Fig. 5). We therefore expect that intrachromosomal recombinations will also occur between g1/g2, r1/r3 and r2/r4 amplicons. These partial AZFc deletions will not delete all gene copies of the AZFc gene families but only part of them (Fig. 5).

All homologous AZFc amplicons have a surprisingly high sequence homology (>99.9%) which could be only distinguished by small sequence nucleotide variants (SNVs) and some sequence tagged sites (STSs). Such a high sequence homology suggests frequent sequence alignments and gene conversions between them. The formation of palindromic structures during some replication of the AZFc DNA sequence would mediate these alignments by the inverted orientation of the duplicated 135 kb sequence block (r1/r2) in proximal AZFc (P2 palindrome) and the duplicated and inverted 1.5 Mb se-

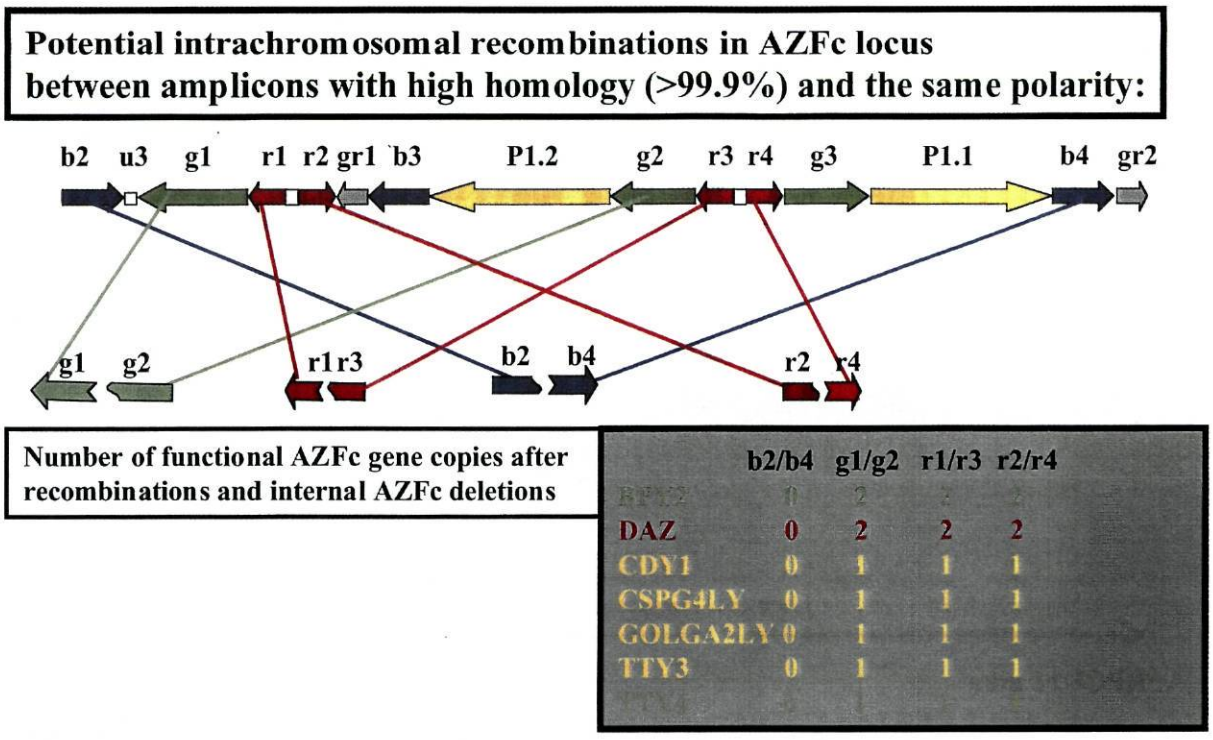


Fig. 5. Potential deletions in the AZFc locus due to intrachromosomal recombinations of four different amplicons organised in AZFc with the same polarity. The number of gene copies of each of the AZFc genes which are still present after occurrence of one of these AZFc deletions is listed on the right. Their colour code correpond to that of the AZFc amplicons where they were mapped by Kawaguchi-Kuroda et al. (2001). Note that only after deletion of the complete AZFc sequence (b2/b4 recombination) all AZFc genes are deleted. For further details see text.

quence block in the distal AZFc sequence (P1 palindrome) (Fig. 4B).

POLYMORPHIC DAZ GENE STRUCTURE

The DYS1 polymorphic DNA region in AZFc (DYS1) is now known to be part of the *DAZ* gene locus. The *DAZ* gene family was transposed to the Y chromosome after divergence of New World Monkeys and Old World Monkeys, i.e. during primate evolution 35 million years ago (18, 19). It has been shown to originate from *DAZL1* located on the short arm of chromosome 3 (3p25). Sequence analyses of the DYS1 cognate probe 49f, a 2.5 kb genomic *Eco*-RI fragment of the Y chromosome has revealed that this fragment contains the exons 4, 5, 6 and 7 of the *DAZ1* gene copy (GenBank accesssion no. 414183). After restriction of genomic male DNA with *Taq*I or *Pvu*II and hybridisation with the 49f probe, a complex repetitive pattern of cross hybridising Y fragments was obtained (33, 34). Today we know that these lengths variabilities are most likely due to a variable number of exon 7 copies in *DAZ2*, *DAZ3*, and *DAZ4*. When we compared structures of each *DAZ* gene copy, mapped in the 7 *DAZ*-BAC clones screened from the GenBank database (RPCI-11 BAC donor) (24) with the homologous structures isolated from the BAC contig of two other male donors (CTA and CTB) (35) only the *DAZ1* gene structure was conserved (Fig. 6). It contains 3 tandem repeats of 10.8 kb including exon 2 till exon 6, encoding the functional RNA-Recognition Motif, RRM, besides a tandem repetitive structure of 2.4 kb with nine exon 7 copies. According to the nomenclature of Yen et al. (36) the sequence variants of exon 7 were designated as *A, B, C, D, E, F, X, Y, Z* with the Y variant as the most prominent consensus sequence. Eight variants were found in *DAZ1* and organised in the order: *5′-B-C-D-E-F-E-X-Y-Z-3′* (Fig. 6). Typical for *DAZ2* is the presence of the *A* variant and ab-

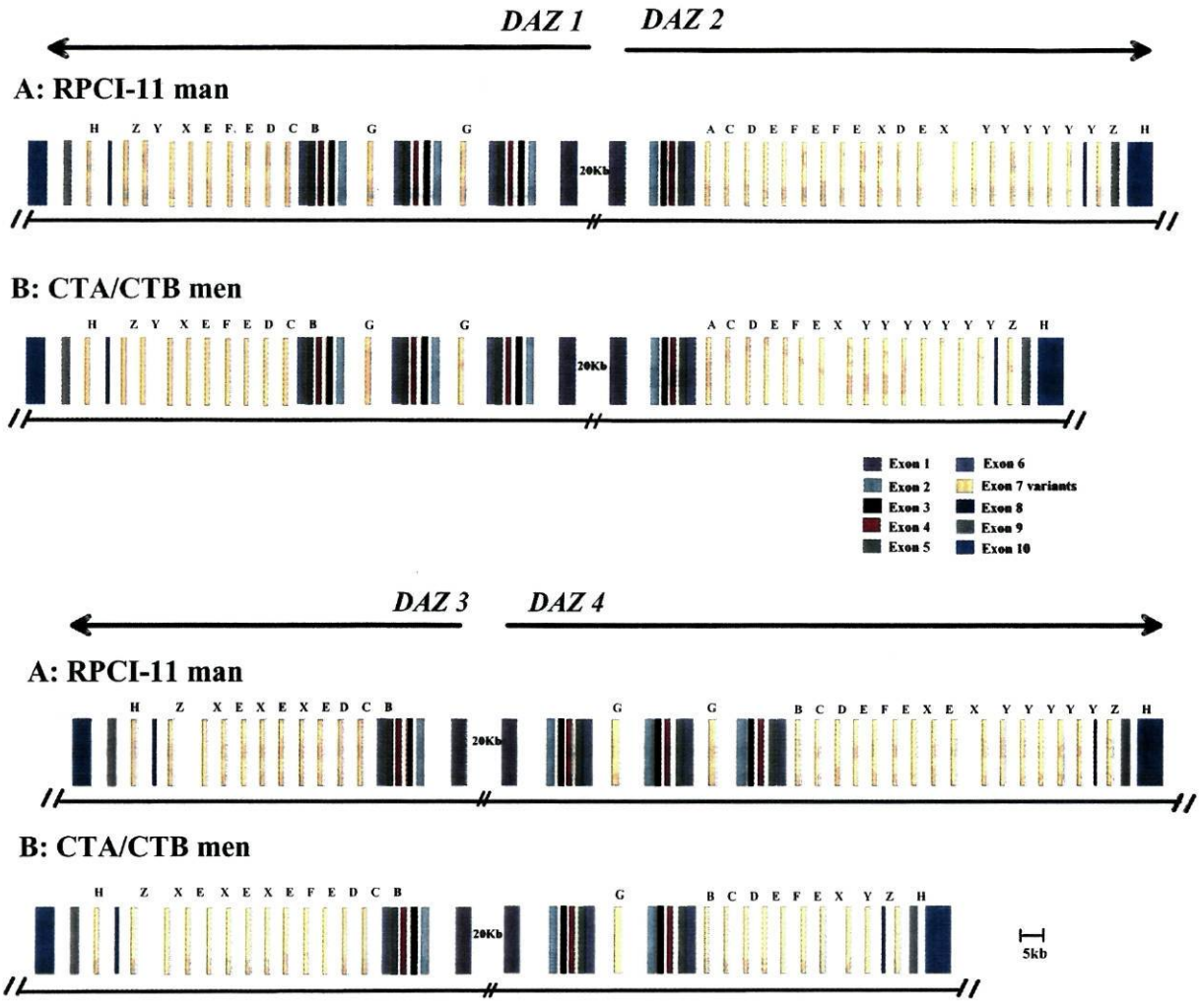


Fig. 6. Detailed view on the exonstructures of the two *DAZ*-doublets: *DAZ1/DAZ2* and *DAZ3/DAZ4* identified in the Y chromosome of the RPCI-11 man (BAC donor for GenBank BAC sequences) and the CTA/CTB men (BAC donors for CTA and CTAB BAC libaries; Saxena et al. 2000) for inspection of their different exon structures described in detail in the text. The letters above the exon 7 variants correspond to the nomenclature given by Yen et al. (1997) with exception of the *G* and *H* variants. These exons were not found in testicular *DAZ* transcripts. The contigs are drawn to scale with a 20 kb gap in the *DAZ* promoter regions.

sence of the *B* variant in the repetitive exon 7 structure besides 6 tandem *Y* variants (Fig. 6). *DAZ2* transcripts with this exon 7 variant structure have been found in testicular *DAZ* cDNA sequences. However, no *DAZ2* testis transcripts were reported containing a duplicated *E–F* variant motif as would be predicted by the genomic *DAZ2* sequence in the RPCI-11 BAC donor. Such *DAZ2* transcripts might therefore be rare polymorphisms or do not exist. The structure of the *DAZ2*-BAC sequence of the CTB library contains an *A* and 7 tandem *Y* variants in the repetitive exon 7 structure and the duplicated *E-F* motif found in the RPCI-11 *DAZ2* sequence is absent.

Substantial differences were found in the repetitive exon structures of *DAZ3* and *DAZ4* located in the r3/r4 amplicons in the distal AZFc region. Typical for *DAZ3* on the Y chromosome of the CTB-donor was the absence of the *Y* variants and a triplication of the *E-X* motif in the repetitive exon 7 structure (35). This was also found in the *DAZ3* gene of the RPCI-11-donor (Fig. 6). However, a *E–F* motif present in the CTB-*DAZ3* gene was not present in *DAZ3* of the RPCI-11 donor. Since testicular

DAZ cDNA sequences were found with both exon 7 variant compositions (see e.g. 36), we can assume that the repetitive exon7 region of the *DAZ3* gene structure is polymorphic on different Y chromosomes and, therefore, a part of the polymorphic DYS1 locus. Another part is the *DAZ4* gene structure which is substantially different in the Y chromosome of the RPCI-11-donor and of the CTB-donor. Both *DAZ4* genes are recognised by their 10.8 kb repeat unit (RRM-repeat), however, three copies of this RRM repeat were found in the RPCI-11-*DAZ4* gene and only two RRM repeats in the CTB-*DAZ4* gene (Fig. 6). The number and composition of the exon 7 variants in both *DAZ4* genes were also different. In the RPCI-11–*DAZ4* gene we found 15 exon 7 copies with the composition: 5′-*B-C-D-E-F-E-X-E-X-Y-Y-Y-Y-Y-Z*-3′ whereas only 9 exon 7 repeats with the composition: 5′-*B-C-D-E-F-E-X-Y-Z*-3′ were identified in the *DAZ4* gene of the CTB-BACs (Fig. 6).

DELETION OF SINGLE DAZ GENE COPIES

The polymorphic *DAZ* gene structures suggest that not all *DAZ* gene copies might have the same functional capacity for human spermatogenesis and that the deletion of single *DAZ* gene copies as a familiar polymorphic deletion event in men with normal fertility can be expected. The *DAZ* gene locus restriction maps and sequence analyses have shown that the genomic structures of the 4 *DAZ* genes were more than 99% homologous to each other. Therefore, before we could analyse putative DAZ gene copy deletions, we had to establish 6 *DAZ*-Single-Nucleotide-Variants (SNV I–VI) spanning the 5′ end till the 3′ end of these *DAZ* gene sequences and 2 *DAZ* gene copy specific STSs (*DAZ*-RRM3, Y-*DAZ3*). The various deletion patterns of these SNV/STS markers distinguish each type of *DAZ* gene and defined deletion of one of the possible 13 *DAZ* deletion haplotypes (Table 1). On the Y chromosome of men with normal fertility, the presence of 4 *DAZ* gene copies is indicated by the presence of the A and B alleles for each of the 6 *DAZ*-SNVs as well as the STSs *DAZ*-RRM3 and Y-*DAZ3*. Absence of all *DAZ* markers would indicate the absence of the complete *DAZ* gene family on the Y chromosome.

When we analysed the presence of one of the 13 *DAZ*-deletion-haplotypes of Table 1 in the genomic DNA samples of 63 men with severe oligozoospermia (OZ samples) and of 107 men with proven fertility (FM samples) surprisingly, 3 *DAZ*-deletion haplotypes were found in both men groups with similar frequencies (*DAZ*-del-hap: 2, 4, 9) (24). These variabilities are therefore most likely polymorphic familiar variants in the *DAZ* gene locus of the Y chromosome of all men of the patients' pedigree. Interestingly, *DAZ*-del-hap 2, was always associated with absence of the B allele for SNV VI. Most likely, this is an additional gene conversion event for SNV VI in the *DAZ* locus of this men group, although a deletion of the distal part of *DAZ4* including the SNV VI *Afl*III restriction site can not be excluded without appropriate DNA blot

TABLE 1. *13* DAZ-*deletion haplotypes are specified by 8* DAZ *SNV/STS markers*

DAZ gene deletions	SNV I	SNV II	SNV III	SNV IV	SNV V	SNV VI	*DAZ*-RRM3	Y-*DAZ3*	*DAZ* del. haplotype
DAZ1	A+B	B	A+B	A+B	A+B	A+B	+	+	1
DAZ2	A+B	A+B	B	B	A+B	A+B	+	+	2
DAZ3	A+B	A+B	A+B	A+B	A+B	A+B	+	−	3
DAZ4	A	A+B	A+B	A+B	A+B	A	+	+	4
DAZ1+DAZ2	A+B	B	B	B	A	A+B	+	+	5
DAZ1+DAZ3	A+B	B	A+B	A+B	A+B	A+B	+	−	6
DAZ1+DAZ4	A	B	A+B	A+B	A+B	A	−	+	7
DAZ2+DAZ3	A+B	A+B	B	B	A+B	A+B	+	−	8
DAZ2+DAZ4	A	A+B	B	B	A+B	A	+	+	9
DAZ3+DAZ4	A	A+B	A+B	A+B	B	A	+	−	10
DAZ1+DAZ2+DAZ3	B	B	B	B	A	B	+	−	11
DAZ1+DAZ2+DAZ4	A	B	B	B	A	A	−	+	12
DAZ2+DAZ3+DAZ4	A	A	B	B	B	A	+	−	13

experiments. Absence of only the A allele of SNV III, or only the A allele of SNV IV (with or without absence of the B allele of SNV VI and absence of the Y-*DAZ3* STS marker) was also identified in both men groups with the *DAZ*-deletion haplotype 2 (24). Although not yet proven, we like to suppose that the most easiest explanation for these variabilities are not local SNV microdeletions but frequent gene conversion events in the inverted r1/r2 amplicons probably mediated by the formation of the P2 palindrome as discussed above. Similar patterns of allelic SNV and STS deletion events were also observed for men with *DAZ*-del hap 4 (indicative for deletion of *DAZ4*) and *DAZ*-del-hap 9 (indicative for deletion of *DAZ2* and *DAZ4*) in both men groups (24). Gene conversion events are probably therefore also mediated by formation of the P2 palindrome in the r3/r4 amplicons. If this holds true, we can take the observed variabilities of the allelic A and B patterns in the SNVs of each *DAZ* gene copy as the first experimental indication that indeed palindromes are formed within the repetitive genomic DNA structure of the AZFc locus.

Contrasting with these results we found a deletion of the *DAZ1* gene linked to a deletion of the *DAZ2* gene (i.e. *DAZ*-deletion-haplotype 5) only in 5 DNA samples of the oligozoospermic men group (24).

DAZ gene copy specific restriction fragments were only found for the DYS1 locus after restricting the genomic DNA samples with *Eco*R V and *Taq* I (24). In this way, a deletion of the *DAZ1* and *DAZ2* genes was confirmed in all samples with *DAZ*-del-hap 5 and we could show, that this deletion is most likely due to an intrachromosomal recombination event between the g1/g2 amplicons in AZFc as suspected in Fig. 5. One copy of the AZFc genes: *BPY2, CDY1, CSPG4LY, GOLGA2LY, TTY3* and *TTY4* should then be included in this partial AZFc deletion (Fig. 5). However, in contrast to the polymorphic structures of the *DAZ* genes, all these AZFc genes have an identical exon structure in each of their gene copies. A deletion of only one of their copies after a g1/g2 based partial AZFc deletion is therefore not expected to contribute significantly to the patients' pathological phenotype, since all copies will have the same functional capacity. A polymorphic structure has been exclusively found in the *DAZ* gene family. We, therefore, assume that in contrast to the other AZFc candidate genes, not all *DAZ* genes might be essential for human spermatogenesis.

SUMMARY

With aid of the detailed molecular picture of the AZFc locus and its gene families it can now be shown that the early observation of polymorphic chromatin domains in the human Y chromosome can not be taken as indicative for a non-functional state of this chromosome or of the Y genes being part of it. However, it is indeed mandatory to study each Y gene mutation identified in AZFc, whether deletion or point mutation, also for its possible presence in the AZFc locus of the patients' father, in the AZFc locus of men with normal fertility, before one associates this AZFc gene mutation with the occurrence of the patients' testicular pathology.

If the proposed structural dynamic arrangements really would occur in the AZFc locus, it is the first spermatogenesis locus in the human genome with a locus-specific chromatin folding code according to our predictions postulated 12 year ago (31, 32). In AZFc this code is based on a unique repetitive sequence structure which is organised in palindromes to stabilise the functional AZFc gene content. Local instabilities usually inherited with tandem repetitive sequence blocks and especially with palindromic sequence structures (37) should become inhibited by this folding code in the AZFc amplicons. This let the AZFc structure look as a structural artwork probably evolved step by step in an ancestral polymorphic heterochromatic Y domain.

In contrast to mouse, rat and bull, human spermatogenesis looks highly disorganized and even healthy and fertile men ejaculate a high percentage of spermatozoa with abnormal morphology and motility. Perhaps this might also explain the variable testicular pathologies found in testis tissue sections of sterile men with AZFc deletions which in rare cases can be inherited. In these familiar AZFc deletions the number and quality of sperms should be still sufficient for the men' fertility. Most likely, the primary effect of the AZFc deletions is therefore only a significant reduction in the number of mature sperms which then leads secondarily to an in-

creased loss of germ cells and a Sertoli-cell-only syndrome (7).

We are grateful to Ewa Rajpert-De Meyts for critical reading this manuscript. We also wish to thank Mrs. Christine Mahrla for their help in preparing the final version of this manuscript. This study was supported by grants to S.F. from the Fundacao para a Ciência e Tecnologia (FCT) (SFRH/BD/811/2000) and to P.H.V. from the Deutsche Forschungsgemeinschaft (DFG: Vo403/10-2).

REFERENCES

1. Unnérus V, Fellman J, De La Chapelle A. The length of the human Y chromosome. Cytogenetics 1967;6:213–27.
2. Mikelsaar AV, Tuur SJ, Kaosaar ME. Human karyotype polymorphism. I. Routine and fluorescence microscopic investigation of chromosomes in a normal adult population. Humangenetik 1973;20:89–101.
3. Neu RL, Matthew MS, Barlow Jr MD, Gardner LI. A 46,XYq- male with azoospermia. Fertil Steril 1973;24:811–3.
4. Tiepolo L, Zuffardi O. Localization of factors controlling spermatogenesis in the nonfluorescent of the human Y chromosome long arm. Hum Genet 1976;34:119–24.
5. Vogt PH. Human Y chromosome function in male germ cell development. Advances in Developmental Biol 1996;4:191–257.
6. Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter, et al. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet 1996;5:933–43.
7. Vogt PH. Human chromosome deletions in Yq11, AZF candicate genes and male infertility: history and update. Hum Reprod 1998;4:739–44.
8. Krausz C, McElreavey K. Y chromosome and male infertility. Front Biosci 1999;4:E1–8.
9. Simoni M. Molecular diagnosis of Y chromosome microdeletions in Europe: state-of-the-art and quality control. Hum Reprod 2001;16:402–9.
10. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC. A physical map of the human Y chromosome. Nature 2001;409:943–5.
11. Lahn BT, Page DC. Functional coherence of the human Y chromosome. Science 1997;278:675–80.
12. Vogt PH, Affara N, Davey P, Hammer M, Jobling MA, Lau YFC, Mitchell M, Schempp W, Tyler-Smith C, Williams G, Yen P, Rappold GA. Report of the third international workshop on Y chromosome mapping 1997. Cytogenet Cell Genet 1997;79:1–20.
13. Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum H, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 2001;29:279–86.
14. Ma K, Inglis JD, Shakey A, Bickmore WA, Hill RE, Prosser EJ, et al. A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. Cell 1993;75:1287–95.
15. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, et al. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 1995;10:383–93.
16. Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC. An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat Genet 1999;23:429–32.
17. Prosser J, Inglis JD, Condie A, Ma K, Kerr S, Thakrar R, Taylor K, Camron JM, Cooke HJ. Degeneracy in human multicopy RBM (YRRM), a candidate spermatogenesis gene. Mamm Genome 1996;7:835–42.
18. Shan Z, Hirschmann P, Seebacher T, Edelmann A, Jauch A, Morell J, Urbitsch P, Vogt H. A *SPGY* copy homologous to the mouse gene *Dazla* and the *Drosophila* gene *boule* is autosomal and expressed only in the human male gonad. Hum Mol Genet 1996;5:2005–11.
19. Seboun E, Barbaux S, Bourgeron T, Nishi S, Agulnik A, Egashira M, et al. Gene sequence, localization, and evolutionary conservation on DAZLA, a candidate male sterility gene. Genomics 1997;41:227–35.
20. Soudek D, Langmuir V, Stewart DJ. Variation in the nonfluorescent segment of long Y chromosome. Humangenetik 1973;18:285–90.
21. Vergnaud G, Page DC, Simmler M-C, Brown L, Rouyer F, Noel B, Botstein D, De La Chapelle A, Weissenbach J. A deletion map of the human Y chromosome based on DNA hybridization. Am J Hum Genet 1986;38:109–24.
22. Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, et al. Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum Mol Genet 1996;5:1767–75.
23. Johnson MD, Tho SP, Behzadian A, McDonough PG. Molecular scanning of Yq11 (interval 6) in men with Sertoli-cell-only syndrome. Am J Obstet Gynecol 1989;161:1732–7.
24. Fernandes S, Huellen K, Goncalves J, Dukal H,

Zeisler J, Rajpert De Meyts E, Skakkebaek NE, Habermann B, Krause W, Sousa M, Barros A, Vogt PH. High frequency of *DAZ1/DAZ2* gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 2002;8:286–98.
25. Horvath JE, Bailey JA, Locke DP, Eichler EE. Lessons from the human genome: transitions between euchromatin and heterochromatin. Hum Mol Genet 2001;10:2215–23.
26. Bühler EM. A synopsis of the human Y chromosome. Hum Genet 1980;55:145–75.
27. Jalal SM, Pfeiffer RA, Pathak S, Hsu TC. Subdivision of the human Y chromosome. Humangenetik 1974;24:59–65.
28. Fernández-Piqueras J, Ludefia P, De Cabo SF, Velázquez M, Sentis C. Longitudinal differentiation of the human Yq heterochromatin as revealed by the restiction enzyme *Taq*I. Cytogenet. Cell Genet 1991;57:78–81.
29. Limon J, Gibas Z, Kaluzewski B, Moruzgala T. Demonstration of two different regions of lateral asymmetry in human Y chromosome. Hum Genet 1979;51:247–52.
30. Bonaccorsi S, Lohe A. Fine mapping of satellite DNA sequences along the Y chromosome of Drosophila melanogaster: relationship between satellite sequences and fertility factors. Genetics 1991;129:177–89.
31. Vogt P. Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". Hum Genet 1990;84:301–36.
32. Vogt P. Code domains in tandem repetitive DNA sequence structures. Chromosoma 1992;101:585–9.
33. Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J. A DNA probe detecting multiple haplotypes of the human Y chromosome. Am J Hum Genet 1986;38:407–18.
34. Spurdle A, Jenkins T. Y chromosome probe p49a detects complex PvuII haplotypes and many new TaqI haplotypes in southern african populations. Am J Hum Genet 1992;50:107–25.
35. Saxena R, de Vries JWA, Repping S, Alagappan RK, Skaletsky H, Brown LG, Ma P, Chen E, Hoovers JMN, Page DC. Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. Genomics 2000;67:256–67.
36. Yen PH, Chai NN, Salido EC. The human DAZ genes, a putative male infertility factor on the Y chromosome, are highly polymorphic in the DAZ repeat regions. Mamm Genome 1997;8:756–9.
37. Akgun E, Zahn J, Baumes S, Brown G, Liang F, Romanienko PJ, Lewis S, Jasin M. Palindrome resolution and recombination in the mammalian germ line. Mol Cell Biol 1997;17:5559–70.

COMMENTS

J. Wolter Oosterhuis (Rotterdam, The Netherlands): What is your take home message which has relevance for a histopathologist studying germ cell tumour biology?

Peter Vogt (Heidelberg, Germany): Not every molecular pathological event results in histological pathology. The polymorphism at the AZFc gene locus indicates that no single deletion will necessarily result in an identifiable disease process in the individual affected. Polymorphic length variations in AZFc are balanced by the presence of multiple copies of the AZFc genes. A pathological phenotype only develops after loss of all gene copies.

Ken McElreavey (Paris, France): How can you distinguish between a true partial deletion of the AZFc gene, and a gene conversion event?

Peter Vogt: Deletions and conversions can only be distinguished using specific Y-DNA probes for genomic blot hybridisation. This can detect a partial deletion it if the probe sequence overlaps a specific restriction fragment which is only present in the AZFc region. For example, DAZ1 is the only gene with a 10.8 kb *Eco*RV genomic fragment of the DYS 1 locus (Fernandes et al. 2002, Mol Human Reprod Vol.3. 286–98).

Ragnhild Lothe (Oslo, Norway): How are the CpG sites in the Y chromosome genes distributed? Do you think that abnormal methylation may be related to their expression in germ cell tumours?

Peter Vogt: Some gene families are clustered in one of the AZF regions, others are spread along the long and short Y chromosome arms. We do not yet know whether methylation processes are involved in the regulation of the expression of the AZF genes.

A large AZFc deletion removes *DAZ3/DAZ4* and nearby genes from Y-haplogroup N men

S. Fernandes[1,2], S. Paracchini[3]*, L. H. Meyer[1], G. Floridia[3], C. Tyler-Smith[3]*** , P. H. Vogt[1]******

Running title: Large AZFc deletion in Y haplogroup N men

[1]Section of Molecular Genetics & Infertility, Department of Gynecological Endocrinology & Reproductive Medicine, University of Heidelberg, Heidelberg, Germany.

[2] Department of Human Genetics, Faculty of Medicine, University of Porto, Porto, Portugal.

[3] Department of Biochemistry, University of Oxford, Oxford, UK.

Present addresses:

*Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

**Laboratorio di Biologia cellulare, Istituto Superiore di Sanità, Roma, Italy.

***The Wellcome Trust Sanger Institute, Hinxton, UK.

******address for correspondence and reprint requests:**

Voßstrasse 9

D-69115 Heidelberg, Germany

Phone: +49-6221-567918 –567916 ; FAX : +49-6221-5633710

E-MAIL: peter_vogt@med.uni-heidelberg.de

Abstract

Deletion of the entire AZFc locus on the human Y chromosome leads to male infertility. The functional roles of the individual gene families mapped to AZFc are, however, still poorly understood since the analysis of the region is complicated by its repeated structure. We have therefore used Single Nucleotide Variants (SNVs) across ~3 Mb of the AZFc sequence to identify 15 AZFc haplotypes and examined them for deletion of individual AZFc gene copies. We found five individuals who lacked SNVs from a large segment of DNA containing the DAZ3/DAZ4 and BPY2.2/BPY2.3 gene doublets in distal AZFc. Southern blot analyses showed that the lack of these SNVs was due to deletion of the underlying DNA segment. Typing 118 binary Y markers showed that all five individuals belonged to Y haplogroup N, and 15/15 independently-ascertained haplogroup N men carried a similar deletion. Haplogroup N is known to be common and widespread in Europe and Asia and there is no indication of reduced fertility in men with this Y chromosome. We therefore conclude that a common variant of the human Y chromosome lacks the DAZ3/DAZ4 and BPY2.2/BPY2.3 doublets in distal AZFc and thus that these genes cannot be required for male fertility: the gene content of the AZFc locus is likely to be genetically redundant. Furthermore, the observed deletions cannot be derived from the GenBank reference sequence by a single recombination event: an origin by homologous recombination from such a sequence organization must be preceded by an inversion event. These data confirm the expectation that the human Y chromosome sequence and gene complement may differ substantially between individuals and more variations are to be expected in different Y-chromosomal haplogroups.

AZFc deletions in distal Yq11 are the most frequent known genetic cause of human male infertility, leading to azoospermia or severe oligozoospermia. They occur in different populations with a frequency of 10-20% of infertile men (Vogt 1998; Krausz and McElreavey 1999; Krausz et al. 2000; Simoni 2001). AZFc contains eight gene families expressed only in testis tissue, three of which code for proteins, but the importance of each individual gene for fertility is not understood. The underlying AZFc sequence is composed mainly of large repeated sequence blocks called amplicons (Kuroda-Kawaguchi et al. 2001) organized into palindromic structures showing high (>99.9%) sequence identity between the arms (see fig. 1). Such structures may undergo frequent inversion and gene conversion events (Skaletsky et al. 2003), as well as duplications and deletions (Repping et al. 2003). We therefore set out to investigate the structure and gene content of this region in normal individuals, using SNVs (Single Nucleotide Variants) and STSs extending along 3 Mb of the AZFc sequence, supplemented by Southern blot analyses.

The assays for analyses of specific variants within the DAZ genes [MIM 400003, 400026, 400027] were described recently (Fernandes et al. 2002): three STSs (DAZ-RRM3 and sY152 in DAZ1 and DAZ4; Y-DAZ3 in DAZ3) and six DAZ-SNVs (I-VI). Typing the SNVs involves coamplification of two or more similar loci followed by restriction enzyme digestion to reveal the absence (A-allele) or presence (B-allele) of a given restriction enzyme site. Since the locations of the A and B-SNV alleles are known from the sequence, loss of the A and/or B pattern indicates a loss of specific DAZ gene copies. Some markers outside the DAZ gene locus were already available: 50f2/C or DYS7/C (GenBank accession no. Y07728) and sY1192 (GenBank accession no. G67166), which both lie in the proximally-located u3 block; and AZFc-P1, an SNV marker which distinguishes the yellow P1.1 amplicon from the P1.2 amplicon (table 1). In addition, we used the AZFc sequence data published in GenBank

to look for SNVs in the seven other gene families within this region: BPY2 (Basic Protein Y2 [MIM 400013]), CDY1 (Chromo-Domain Y1 [MIM 400016]) CSPG4LY (Chondrotin-sulfate proteogylcan-4-like Y), GOLGA2LY (Golgi-Antigene 2 Like Y), TTY3, TTY4, and TTY17 (Testis-Transcript Y 3, 4, 17). We found no suitable variants in the CDY1, CSPG4LY, TTY3 and TTY17 genes located in the yellow and green amplicons P1.1/P1.2 and g1-3, respectively. However, SNVs were found for the genes BPY2 and TTY4 located in the green amplicons (g1-3), and for GOLGA2LY located in the yellow (P1.1/ P1.2) amplicons, and assays were established (table 1). These markers, together with the DAZ-STS/SNVs, provided assays for 37 positions in eight loci spread over ~3000 kb in the AZFc region (fig. 1).

With this marker set, we analyzed the AZFc region in 31 individuals (In1-In31). Fifteen different 'AZFc haplotypes' were present, 14 of which showed loss of variants in our set (table 2). In most cases, one or a small number of variants were missing, and the missing loci were not located adjacent to one another on the physical map (fig. 1). These were interpreted as likely to have arisen in multiple ways such as mutations in the restriction enzyme or primer binding sites, gene conversion events, or small deletions, but do not seem to involve the loss of large segments of DNA. In contrast, two AZFc region haplotypes, containing a total of five individuals (16%), stood out. They lacked the two STSs from u3 and also all SNVs from a large contiguous region in distal AZFc extending from P1.2 and g2, through DAZ3 and DAZ4 to g3 (fig. 1) suggesting loss of a large AZFc segment in distal Yq11.

Local gene conversions or other changes involving small numbers of nucleotides in the DAZ locus can be distinguished from large DAZ deletions by Southern blot experiments using probes of DYS1-DNA from the DAZ gene (Fernandes et al. 2002). Despite the

similarity of the DAZ gene sequences, *Eco*RV or *Taq*I digests produce patterns in which fragments specific for the DAZ gene copies can be distinguished. Similar analyses on a larger scale can be also performed after genomic *Sfi*I restriction digestion. Scoring the DAZ-BAC-contig in GenBank for rare-cutter restriction sites identified ~450 kb and ~350 kb genomic *Sfi*I fragments carrying the DAZ1/DAZ2 gene doublet and DAZ3/DAZ4 gene doublet, respectively (Floridia et al. 2000). We therefore performed such experiments to investigate whether men lacking the SNVs carry a deletion, and the results are shown in Figure 2. A deletion of the DAZ3 and DAZ4 genes of In29 and In30 is indicated by absence of the DAZ3-specific 19.6 kb-*Taq*I fragment and of the DAZ4-specific 7.3 kb-*Eco*RV fragment (Figure 2A). Deletion of the DYS1-*Sfi*I fragment specific for the DAZ3/DAZ4 gene doublet (~350 kb) is seen in Figure 2B for sample In37. Southern blot hybridization therefore shows that substantial DNA segments are missing, and we conclude that the absence of the P1.2-g2-DAZ3-DAZ4-g3 markers in the Y chromosome of these individuals is due to a large DNA deletion in this part of the AZFc region. We therefore refer to these chromosomes as carrying a DAZ3/DAZ4 deletion.

In parallel, the same DNA samples were typed with 118 Y-chromosomal binary markers to establish their haplogroup (Paracchini et al. 2002). Results are given according to the Underhill-haplogroup numbering (h1-h116; (Underhill et al. 2000) and the current haplogroup nomenclature of the Y Chromosome Consortium (Y-Chromosome-Consortium 2002) (table 2, fig. 3). Only 10 of the 116 haplogroups found worldwide were present in our samples, as would be expected from their predominant origin from Europe. Eight of the 10 haplogroups are common in Europe; the other two (O3* and Q3*), found in one individual each, are common in the countries of origin of these individuals, China and Costa Rica, respectively. There were striking (although not total) correspondences between the AZFc

haplotypes and the Y-SNP phylogeny (table 2) and some of the AZFc variants could be placed on the Y phylogeny as unique mutations while others were recurrent. The DAZ3/DAZ4 deletion chromosomes all belong to haplogroup N3* and thus suggest a single origin for this AZFc deletion.

To investigate the relationship between this deletion polymorphism and the phylogeny further, we carried out the reciprocal study, testing 15 individuals (In32-In46), who had initially been ascertained as belonging to the N lineage, for their AZFc deletion haplotype (table 2). There was some variation in DAZ-SNV II, but, most strikingly, all except one had lost both u3 and the same P1.2-g2-DAZ3-DAZ4-g3 markers lacking from the five haplogroup N individuals in the first sample set. The amplification of sY1192 in the one exceptional individual is most likely due to cross-reaction with the related Yp sequence block (Skaletsky et al. 2003). We therefore conclude that a common Y lineage, haplogroup N, lacking numerous AZFc markers, carries a deletion of two segments of the GenBank AZFc reference sequence.

Intrachromosomal deletions commonly arise by homologous recombination between repetitive sequence blocks lying in the same orientation (Chen et al. 1997). For example, complete AZFc deletions were accounted for by intrachromosomal recombination between the b2 and b4 amplicons (Kuroda-Kawaguchi et al. 2001) and, similarly, deletions of the DAZ1/DAZ2 gene doublet were accounted for by intrachromosomal recombination between the g1 and g2 amplicons (Fernandes et al. 2002). We therefore investigated whether the lineage N DAZ3/DAZ4 deletions could be explained in a similar way. There were, however, no obvious candidates for directly repeated amplicons flanking the two deleted regions of the N lineage Y chromosome (fig. 1). If the deletion substrate in the pre-N-lineage Y

chromosome resembled the GenBank sequence, two independent deletions must have occurred, neither based on long regions of homology, but one removing u3 and the other the P1.2-g2-DAZ3-DAZ4-g3 region. The deletion, however, was an ancient event (see below) and did not occur in a modern sequence. The deletion substrate could therefore differ from the GenBank AZFc sequence with an organization in the Y chromosomes of the pre-N lineage, that carries u3 adjacent to the P1.2-g2-DAZ3-DAZ4-g3 region, so that a single deletion event would remove all of the sequence missing in the N lineage by an intrachromosomal recombination event between homologous amplicons similar to that seen for complete AZFc deletions with the GenBank AZFc sequence.

The GenBank sequence, which is largely derived from the RPCI-11 donor (Kuroda-Kawaguchi et al. 2001), belongs to Y chromosome haplogroup R because it carries the derived state of the marker M207 (position 139,206 in the RPCI-11 BAC clone 386L3; GenBank accession no. AC006376). The finding that the AZFc-P1 genotype A+B, present in the RPCI-11 donor, is restricted to haplogroup R1* (table 2) also supports this assignment. Haplogroups N and R are not closely related, and their most recent common ancestor probably lived around 36,000 ± 6,000 years ago (Hammer and Zegura 2002), so substantial differences in structure, including inversions, could have accumulated especially in the large palindromic structure in distal AZFc where homologous recombination could lead to inversions of the intervening DNA (Zhou et al. 2001). A rearrangement of the GenBank AZFc structure could place u3 adjacent to the P1.2-g2-DAZ3-DAZ4-g3 region: inversion between the b2 and b3 amplicons (fig. 4). Then a g1/g3 recombination could delete all the sequences found to be absent from lineage N chromosomes in a single step (fig. 4). If this holds true, we must consider that the GenBank AZFc amplicon structure of the R-lineage Y chromosome is only one of many possible arrangements, and that proposed here for the AZFc structure in the

N-lineage Y chromosome is only one other example. Indeed, the large P1 palindrome in distal AZFc was proposed to originate by duplication and inversion of an ancestral AZFc sequence similar to that proposed here for the pre-N lineage AZFc sequence (Kuroda-Kawaguchi et al. 2001).

Y chromosomes carrying a large AZFc deletion variant removing the DAZ3 and DAZ4 genes may be common. In our sample, they corresponded precisely to haplogroup N chromosomes, including members of both the lineages N3* and N*(xN3). It therefore seems likely that all haplogroup N chromosomes carry this deletion. If so, we can use the known distribution of haplogroup N to deduce their frequency and distribution. This haplogroup is common throughout northern Europe and Asia, making up ~12% of Y chromosomes in one worldwide survey and forming the majority in some populations such as the Finns (~52%) and Yakuts (Sakha, ~86%) (Zerjal et al. 1997). It forms a subset of the chromosomes previously identified as lacking 50f2/C (Jobling et al. 1996) but was erroneously considered to carry intact DAZ genes in this early study. The most recent common ancestor of haplogroup N chromosomes is estimated to have lived about 8,800 ± 3,200 years ago (Hammer and Zegura 2002). It is therefore an ancient and successful lineage, and the lack of DAZ3 and DAZ4 has no detrimental effect on these men's fertility. Other large Y deletion variants are for example a ~3 Mb segment of Yp containing the AMEL and PRKY genes which also appears to be neutral as it was absent from some normal males (Santos et al. 1998). Another partial AZFc deletion designated the "gr/gr" deletion seems however not to be neutral since it was described as a significant risk factor for spermatogenic failure (Repping et al. 2003). This AZFc structure can arise by recombination between g1 and g2, but also between r1 and r3, or r2 and r4, respectively, in the AZFc GenBank sequence. No inversions are needed to arrange these amplicons into the same polarity (fig. 4). All gr/gr deletions

therefore create an AZFc structure which is different from that proposed here for Y-haplogroup N men. This is further illustrated by the gr/gr deletion's retention of u3 but lack of the r1, r2, b3 and gr1 amplicons; in contrast, in Y haplogroup N u3 is deleted and r1, r2, b3 and gr1 are present with reverse polarity (fig. 4). Interestingly, gr/gr deletions were not reported in men with Y-haplogroup N, although they were found in 14 other Y haplogroups (Repping et al. 2003). AZFc gr/gr deletions therefore appear to have arisen on many occasions and be continually eliminated by natural selection because of their modest negative impact on men's fertility (Repping et al. 2003) while the g1/g3 deletion described here seems to be associated only with Y-haplogroup N. It is not known whether the gr/gr deletion underlies the reduced sperm count associated with one haplogroup in Danish males (Krausz et al. 2001), or includes the g1/g2 deletions found in the AZFc structure of some men with oligozoospermia and deletion of the DAZ1/DAZ2 gene doublet (Fernandes et al. 2002). If not, yet more deletions may be found in the future.

If the proposed g1/g3 recombination event is the origin of the AZFc deletion in the N-lineage Y chromosomes, the length of the AZFc sequence in these men is reduced by more than 50% (~3.7 vs. ~1.5 Mb) and with this two copies of the BPY2 gene family and also one copy each of the probably non-coding GOLGA2LY and TTY4 gene families are deleted (fig. 4). The functional contribution of the AZFc gene families to human spermatogenesis therefore appears to be genetically redundant and the roles of the individual genes may be most readily understood by investigating mutations in populations like the Finns where many men carry fewer AZFc genes. Much further work is needed to understand the complete spectrum of normal and pathogenic variations in the structure of the Y-chromosomal AZFc region. The present study and forthcoming ones have implications for interpreting the results of Y STS deletion analyses in the infertility clinic, illustrating the care that must be taken

when deletions are discovered and their phenotypic consequences for male fertility need to be predicted.

Acknowledgements

We thank all the donors of the genomic DNA samples and are grateful to Karin Huellen und Alexandra Schadwinkel for their help in collecting and isolating them from the individual blood samples. We also wish to thank Mrs. Christine Mahrla for her help in preparing the final version of this manuscript. This study was supported by grants to S.F. from the Fundação para a Ciência e Tecnologia (FCT) (SFRH/BD/811/2000) and to P.H.V. from the Deutsche Forschungsgemeinschaft (DFG: Vo403/10-2). S.P and C.T.-S. were supported by the Cancer Research Campaign (now Cancer Research UK).

Electronic-Database Information

Ensembl genome browser, http://www.ensembl.org/ (for the AZFc sequence structure)

GenBank, http://www.ncbi.nlm.nih.gov/ (for accession numbers for data presented here first are as follows:

GOLY-SNV I GenBank STS database: accession number: BV012731

BPY2-SNV I GenBank STS database: accession number: BV012732

TTY4-SNV I GenBank STS database: accession number: BV012733)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for BPY2, CDY1 and DAZ)

References

Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. Nat Genet 17:154-163

Fernandes S, Huellen K, Goncalves J, Dukal H, Zeisler J, Rajpert De Meyts E, Skakkebaek NE, Habermann B, Krause W, Sousa M, Barros A, Vogt PH (2002) High frequency of *DAZ1/DAZ2* gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 8:286-298.

Floridia G, Gimelli G, Zuffardi O, Earnshaw WC, Warburton PE, Tyler-Smith C (2000) A neocentromere in the DAZ region of the human Y chromosome. Chromosoma 109:318-327

Hammer MF, Zegura SL (2002) The human Y chromosome haplogroup tree: nomenclature and phylogeny of its major dicisions. Annu Rev Anthropol 31:303-321

Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sajantila A, Salo PJ, Nakahori Y, Disteche CM, Thangaraj K, Singh L, Crawford MH, Tyler-Smith C (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum Mol Genet 5:1767-1775

Krausz C, McElreavey K (1999) Y chromosome and male infertility. Front Biosci 4:E1-8

Krausz C, Quintana-Murci L, McElreavey K (2000) Prognostic value of Y deletion analysis: what is the clinical prognostic value of Y chromosome microdeletion analysis? Hum Reprod 15:1431-1434

Krausz C, Quintana-Murci L, Meyts ER, Jorgensen N, Jobling MA, Rosser ZH, Skakkebaek NE, McElreavey K (2001) Identification of a Y chromosome haplogroup associated with reduced sperm counts. Hum Mol Genet 10:1873-1877.

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 29:279-286

Paracchini S, Arredi B, Chalk R, Tyler-Smith C (2002) Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. Nucleic Acids Res 30:e27

Repping S, Skaletsky H, Brown L, van Daalen SKM, Korver CM, Pyntikova, Kuroda-Kawaguchi T, de Vries JWA, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Nat Genet online 5 October 2003; doi: 10.1038/ng1250

Santos FR, Pandya A, Tyler-Smith C (1998) Reliability of DNA-based sex tests. Nature Genetics 18:103

Simoni M (2001) Molecular diagnosis of Y chromosome microdeletions in Europe: state-of-the-art and quality control. Hum Reprod 16:402-409

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358-361

Vogt PH (1998) Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. Mol Hum Reprod 4:739-744

Y-Chromosome-Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339-348

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

Zhou ZH, Akgun E, Jasin M (2001) Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. Proc Natl Acad Sci U S A 98:8326-8333

Tables

Table 1: AZFc-SNVs and STSs

SNVs	AZFc site	GenBank accession no.	SNV restriction enzyme	AZFc-amplicon	Ensembl Y-DNA seq. position (bp)	SNV allele	Restricted fragments (bp)
AZFc-P1 SNV I	P1.1/P1.2	G73351		P1.1	26.709.824 - 26.710.291	A	467
			*Dde*I	P1.2	25.556.438 - 25.556.845	B	284+183
GOLY SNV I	GOLGA2LY genes (2 copies)	BV012731	*Hha*I	P1.1	26.763.718 - 26.764.249	A	531
				P1.2	25.501.889 - 25.502.420	B	289+242
BPY2 SNV I	BPY2 genes (3 copies)	BV012732		g2	25.930.869 - 25.931.339	A	470
				g3	26.335.746 - 26.336.216	A	
			*Eco*RV	g1	24.388.898 - 24.389.368	B	289+181
TTY4 SNV I	TTY4 genes (3 copies)	BV012733		g3	26.368.272 - 26.368.812	A	540
			*Hae*III	g1	24.356.372 - 24.356.912	B	323+218
				g2	25.898.343 - 25.898.883	B	

STSs	AZFc site	GenBank accession nr.	BAC clone(s) in GenBank	BAC seq.-position	Ensembl Y-DNA seq. position (bp)	Anneal. temp. ^{0}C	PCR product (bp)
sY1192	u3	G67166	BAC5C5; AC010080	52109-52362	24.019.703 - 24.020.304	64	255
50f2/C	u3	Y07728	BAC5C5; AC010080	65711-66830	24.033.305 - 24.034.424	65	1119

Table 2. Summary of AZFc-STS and SNV analyses and Y-haplogroup typing in 46 genomic DNA samples

DNA-sample	DAZ-SNV I	DAZ-SNV II	DAZ-SNV III	DAZ-SNV IV	DAZ-SNV V	DAZ-SNV VI	Y-DAZ3	50f2/C	sY1192	TTY4-SNV I	BPY2-SNV I	GOLY-SNV I	AZFc-P1-SNV I	AZFc del ht	Y-haplogroup	YCC-code
In1	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In2	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A	A+B	B	2	108	R1a1*
In3	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In4	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In5	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A	A+B	B	2	108	R1a1*
In6	A+B	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	1	104	R1*
In7	A+B	A+B	B	B	A+B	A	+	+	+	A+B	A	A+B	B	3	52	I*
In8	A+B	A+B	B	B	A+B	A	+	+	+	A+B	A	A+B	B	3	52	I*
In9	A+B	A+B	B	B	A+B	A	+	+	+	A+B	A	A+B	B	3	52	I*
In10	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A+B	B	B	4	58	J2*
In11	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A	A+B	B	3	35	E3b1*
In12	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A+B	A+B	B	5	56	J2f*
In13	A+B	A+B	B	B	A+B	A	+	+	+	A	A	A+B	B	6	58	J2*
In14	A+B	A+B	B	A+B	A+B	A	+	+	+	A+B	A	A+B	B	3	32	E3b3a*
In15	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	7	52	I*
In16	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	7	35	E3b1*
In17	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	7	32	E3b3a*
In18	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	7	35	E3b1*
In19	A+B	A+B	A+B	B	A+B	A	-	+	+	A+B	A	A+B	B	3	35	E3b1*
In20	A	A+B	A+B	A+B	A+B	A+B		+	+	A+B	A+B	A+B	A+B	8	104	R1*
In21	A	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	A+B	8	104	R1*
In22	A+B	A+B	A+B	A+B	A+B	A	+	+	+	A+B	A+B	A+B	B	9	115	Q3*
In23	A	A+B	A+B	A+B	A+B	A+B	+	+	+	A+B	A+B	A+B	B	10	108	R1a1*
In24	**A**	**A+B**	**A+B**	**A+B**	**A+B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**11**	**92**	**N3***
In25	**A**	**A+B**	**A+B**	**A+B**	**A+B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**11**	**92**	**N3***
In26	A	A+B	A+B	B	A+B	A	+	+	+	A+B	A+B	A+B	B	12	76	O3*
In27	A	A+B	A+B	B	A+B	A	+	+	+	A+B	A	A+B	B	13	52	I*
In28	A	A+B	B	A+B	A+B	A	+	+	+	A+B	A+B	A	B	14	58	J2*
In29	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**92**	**N3***
In30	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**92**	**N3***
In31	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**92**	**N3***
Summary of AZFc-STS and SNV analyses in 15 genomic DNA samples from Y-haplogroup N																
In32	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**+**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3)**
In33	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3)**
In34	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3)**
In35	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3)**
In36	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3)**
In37	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**92**	**N3***
In38	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**92**	**N3***
In39	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**92**	**N3***

In40	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**92**	**N3***
In41	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3**
In42	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3**
In43	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3**
In44	**A**	**A+B**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**15**	**87**	**N*(xN3**
In45	**A**	**A**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***
In46	**A**	**A**	**A+B**	**A+B**	**B**	**A**	**-**	**-**	**-**	**B**	**B**	**A**	**B**	**16**	**92**	**N3***

Figure Legends

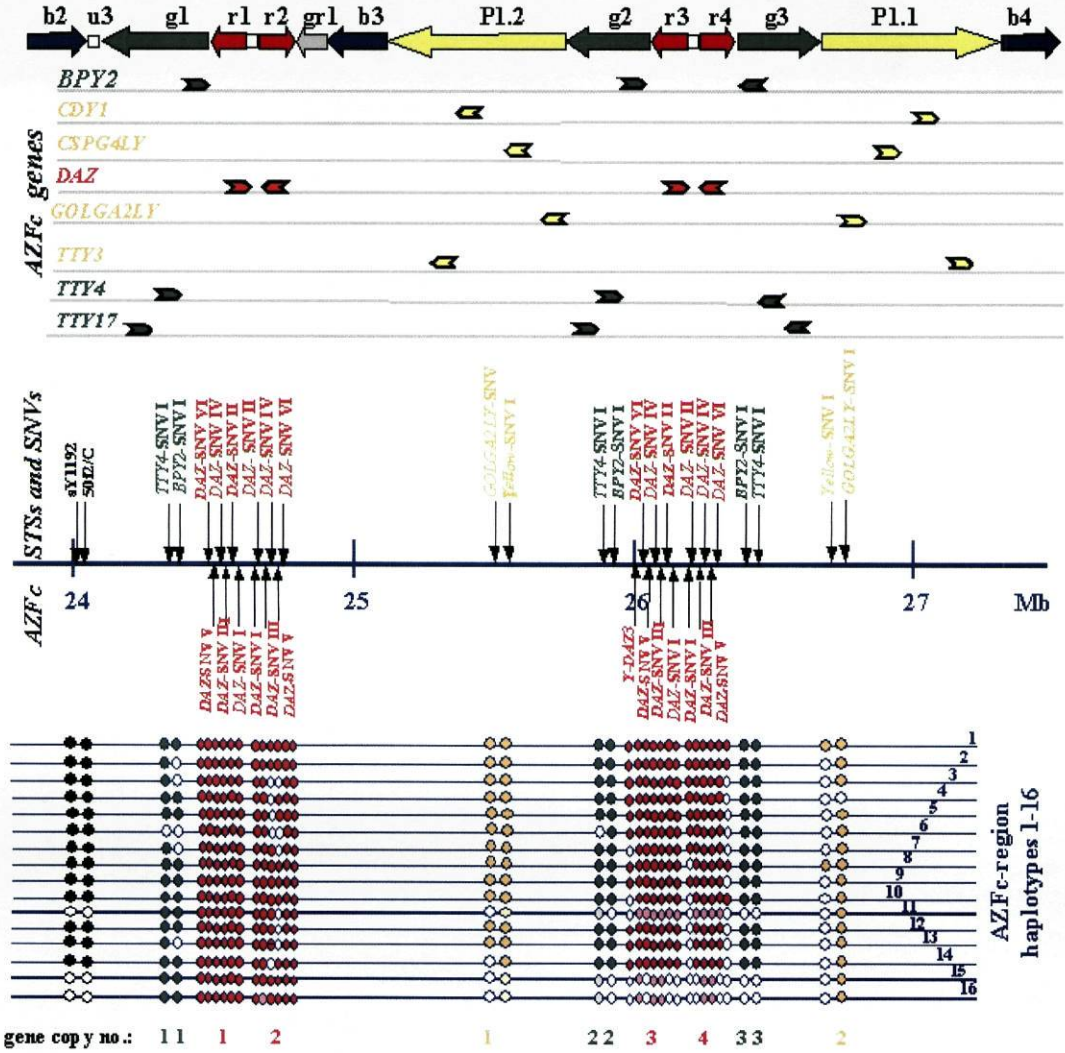
Figure 1. Schematic view of the AZFc locus showing its gene content and the locations of the STSs and SNVs used in this study. The AZFc amplicon structure is drawn according to the color code of Kuroda-Kawaguchi et al. (2001). AZFc genes are listed below the amplicon structure, with their 5′-3′ polarities indicated by directions of the arrows. The STSs and SNVs used in this study are shown below the gene map. Their GenBank accession numbers and molecular data are listed in Fernandes et al. (2002) and table 1, respectively. All AZFc markers and genes were mapped onto the Ensembl Y-chromosome-sequence (http://www.ensembl.org) between 23.79 and 27.49 Mb. Including the complete b2 and b4 amplicons, a length of ~3.7 Mb for the AZFc sequence was estimated. The presence or absence of an STS or SNV in the genomic DNA samples analyzed could be grouped into 16 patterns (AZFc deletion haplotypes 1-16). These are symbolized below the STS/SNV map by rows of filled or empty circles, with the gene copies numbered at the bottom of the figure. Thick lines represent deletion haplotypes belonging to haplogroup N. The pale colors of some circles indicates that these multi-locus SNVs were inferred to be absent because all unambiguously-typed flanking and intervening markers were absent.

Figure 2. **A**: *Eco*RV and *Taq*I DAZ deletion analyses. According to Fernandes et al. (2002) the DAZ3/DAZ4 deletion in the genomic DNA samples of In29 and In30 is shown by absence of the *DAZ3*-specific 19.6 kb *Taq*I fragment and the DAZ4-specific 7.3 kb *Eco*RV fragment. Some genomic DYS1-*Taq*I fragments only visible in the male controls (9.6 kb, 5 kb, 4 kb) are common in all DAZ genes but reduced in intensity in the In39 and In30 samples because of the DAZ3/DAZ4 deletion. The length of the 19.6 kb *Taq*I fragment in the DAZ3 gene is known to be polymorphic, as shown here for the male control samples Mcon1 (with a 19.6 kb *Taq*I fragment) and Mcon2 (with a 24 kb *Taq*I fragment) in the DAZ3 gene. **B.** *Sfi*I DAZ-blot deletion pattern of the genomic DNA sample of In37. The *Sfi*I site at the proximal end of each fragment is partially resistant to cleavage (Floridia et al. 2000), so pulsed-field gel electrophoresis produces a pattern in which hybridization is also seen to partially-cleaved products containing the adjacent ~196 kb fragment: ~649 kb and ~543 kb respectively. Both *Sfi*I fragments specific for the DAZ3/DAZ4 gene doublet (~350 and ~550 kb) are deleted in In37, confirming the complete loss of these DAZ genes. The genomic male (M) and female (F) control samples display the normal *Sfi*I fragments associated with the two DAZ gene doublets DAZ21/DAZ2 and DAZ3/DAZ4 (M control) and with the autosomal DAZL1 gene on chromosome 3 (~380 kb *Sfi*I fragment) in M and F control sample due to cross-hybridisation. Molecular lengths of the λ-*Hin*dIII (**A**) and λ-multimer (**B**) size markers are given on the left.

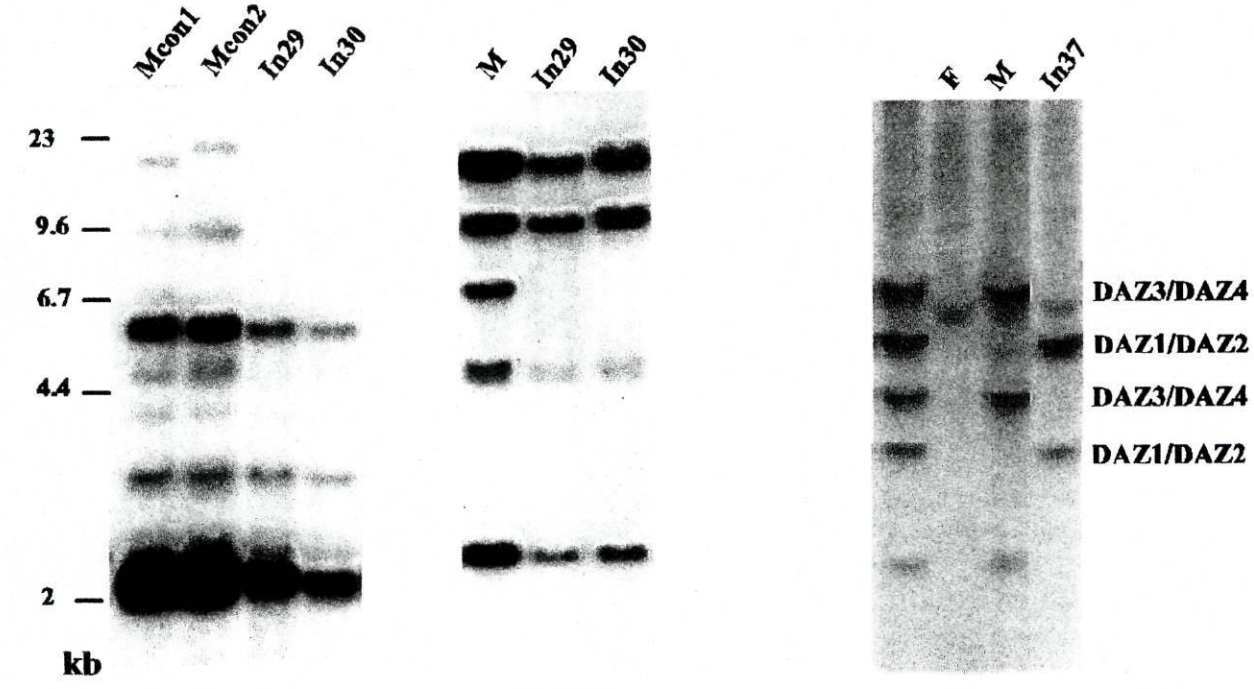
Figure 3. Phylogeny of the Y chromosomes analyzed in this work. Selected Y-SNPs are shown in boxes, and the haplogroup names according to the YCC (2002) or Underhill et al (2000) are shown at the bottom. Only the haplogroups present in the samples analyzed are included. The phylogeny represents the branching order, but not timescale. In accordance with the YCC convention, haplogroups that may not be monophyletic (paragroups) are

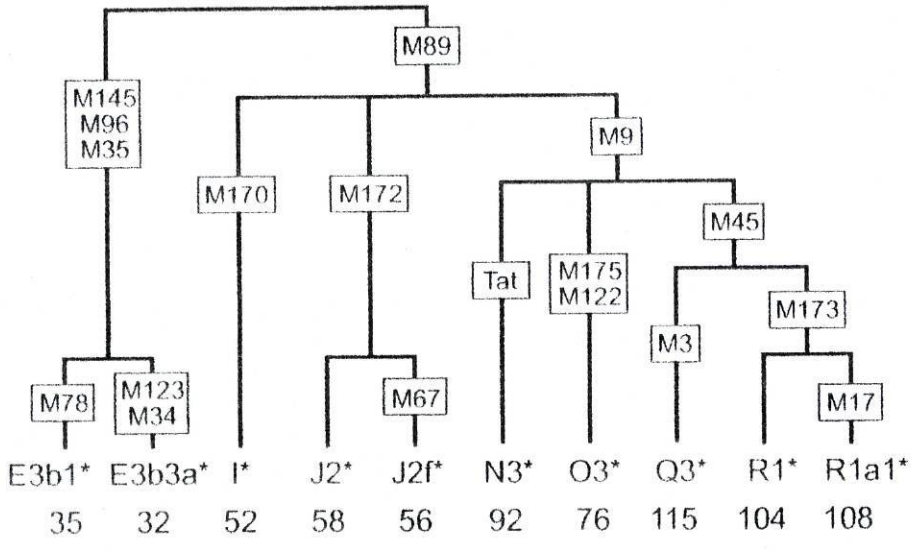
labelled with an *.

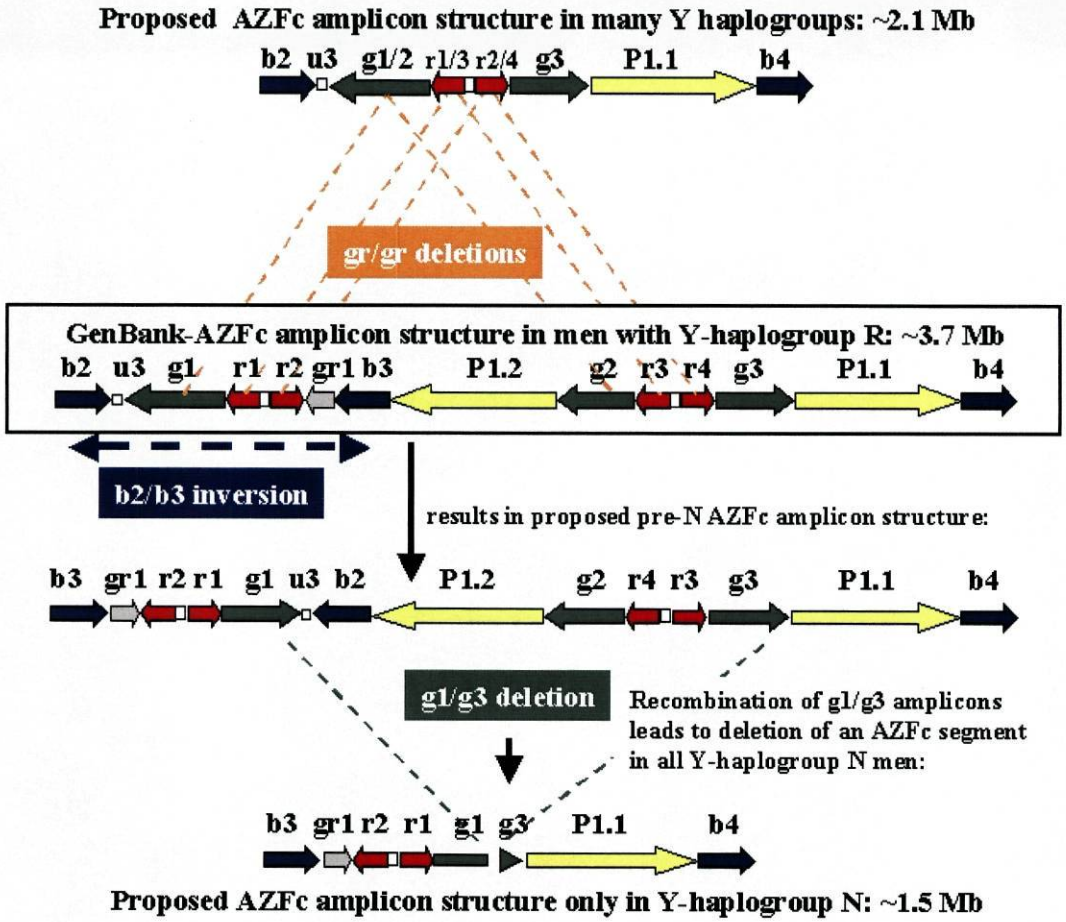
Figure 4. Schematic representation of putative genomic rearrangements in the amplicon structure of the AZFc sequence in men with Y-haplogroup R (GenBank reference sequence) that could lead to an AZFc amplicon structure with u3 adjacent to the P1.2-g2-DAZ3-DAZ4-g3 AZFc segment. After the proposed inversion, mediated by b2/b3, the g1 and g3 amplicons flank the contiguous u3-P1.2-g2-DAZ4-DAZ3 AZFc segment with direct polarity. Recombination between the g1/g3 amplicons in this proposed pre-N AZFc sequence would then lead to deletion of an AZFc segment in men with Y haplogroup N as indicated. This AZFc structure is different from that proposed by Repping et al. (2003) for structures with gr/gr deletions found in 14 different Y haplogroups but not in N (see above GenBank reference sequence). Recombinations of gr/gr deletions can take place between g1/g2 amplicons, r1/r3 amplicons, and r2/r4 amplicons, respectively, as indicated by the 3 brownish lines but not between the g1/g3 amplicons as found in men with Y haplogroup N. For further details see text.


b2 u3 g1 r1 r2 gr1 b3 P1.2 g2 r3 r4 g3 P1.1 b4
AZFc genes
BPY2
CDY1
CSPG4LY
DAZ
GOLGA2LY
TTY3
TTY4
TTY17
STSs and SNVs
sY1192
50f2/C
TTY4-SNV I
BPY2-SNV I
DAZ-SNV VI
DAZ-SNV IV
DAZ-SNV II
DAZ- SNV II
DAZ-SNV IV
DAZ-SNV VI
GOLGA2LY-SNV
Yellow- SNV I
AZFc
24
25
26
27
Mb
DAZ5 SNV V
DAZ-SNV III
DAZ-SNV I
Y-DAZ5
AZFc-region haplotypes 1-16
gene copy no.: 1 1 1 2 1 2 2 3 4 3 3 2

DAZ3/4 specific restriction fragments deleted in men with haplogroup N




M89
M145
M96
M35
M9
M170
M172
M45
Tat
M175
M122
M173
M3
M78
M123
M34
M67
M17
E3b1*
E3b3a*
I*
J2*
J2f*
N3*
O3*
Q3*
R1*
R1a1*
35
32
52
58
56
92
76
115
104
108


Proposed AZFc amplicon structure in many Y haplogroups: ~2.1 Mb
b2 u3 g1/2 r1/3 r2/4 g3 P1.1 b4
gr/gr deletions
GenBank-AZFc amplicon structure in men with Y-haplogroup R: ~3.7 Mb
b2 u3 g1 r1 r2 gr1 b3 P1.2 g2 r3 r4 g3 P1.1 b4
b2/b3 inversion
results in proposed pre-N AZFc amplicon structure:
b3 gr1 r2 r1 g1 u3 b2 P1.2 g2 r4 r3 g3 P1.1 b4
g1/g3 deletion
Recombination of g1/g3 amplicons leads to deletion of an AZFc segment in all Y-haplogroup N men:
b3 gr1 r2 r1 g1 g3 P1.1 b4
Proposed AZFc amplicon structure only in Y-haplogroup N: ~1.5 Mb