The Human *Deleted in Azoospermia* Gene Family: Structure, Function and Evolution

by

Richa Saxena

B.A. Biological Sciences and Chemistry
Cornell University, 1991

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2000

Signature of Author: _____
Department of Biology
April 28, 2000

Certified by: _____
David C. Page
Professor of Biology
Howard Hughes Medical Institute
Thesis Supervisor

Accepted by: _____
Terry Orr-Weaver
Professor of Biology
Co-chair, Biology Graduate Student Committee

The Human *Deleted in Azoospermia* Gene Family: Structure, Function and Evolution

by

Richa Saxena

Submitted to the Department of Biology on April 28, 2000
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Biology

ABSTRACT

The human *Deleted in Azoospermia* genes lie within the *AZFc* region of the Y chromosome that is frequently deleted in men with severe spermatogenic failure and male infertility. This thesis describes the evolution, structure and function of human *DAZ* genes. First, we determined that Y-linked *DAZ* genes arose by transposition from an autosomal ancestor during primate evolution, and were repeatedly amplified and pruned by multiple splicing mutations. This was the first example of acquisition of a male fertility factor by the Y chromosome independent of the X chromosome in any species. Then, we characterized the genomic structure and organization of four Y-linked *DAZ* genes in two clusters within the *AZFc* region. We found that each cluster has an inverted head to head pair of two *DAZ* genes. All *DAZ* genes appear capable of encoding DAZ proteins and two of these DAZ genes encode proteins with multimeric RNA-binding domains. We found transcripts from three *DAZ* genes. Next we determined the molecular nature of *DAZ* gene polymorphism in men. We deciphered *49f/a Taq*I polymorphism, which corresponds to polymorphism in 2.4 kb repeats within all *DAZ* genes, and studied homogenization and deletion of one or the other *DAZ* gene cluster. We find that *DAZ* genes and corresponding *DAZ* proteins are highly polymorphic in different human Y chromosomal lineages. In the first appendix, we describe characterization of *AZFc* region deletions in three infertile men, two of whom have deletions of one *DAZ* cluster. In the second appendix, we describe partial rescue of the meiotic entry defect in *Drosophila boule* by overexpression of human DAZ and mouse DAZL proteins in testes of boule deficient flies. This functional assay suggests some functional conservation of DAZ and boule function. This thesis uncovers the genomic complexity, dynamic instability and evolution of a Y chromosomal gene family that is important for its role in human male fertility.

Thesis Supervisor: David C. Page
Title: Professor of Biology, Howard Hughes Medical Institute

To Papa, Mummy, Nipun and Sagun

Acknowledgments

I would like to express my deepest appreciation to my dear colleagues and friends in the Page Lab who supported me along the way both professionally and personally. Laura Brown and Raaji Alagappan have been crucial to my work from the beginning and I am very grateful to both of them. Isin Dalkilic suffered through the complexities of DAZ genomics with me, and brought much sunshine into my life. Charles Tilford, Doug Menke and earlier Bruce Lahn have been personal and scientific companions . I have treasured the warm friendship and advice of Alex Bortvin and Tomoko Kawaguchi. I would also like to thank Helen Skaletsky and Steve Rosen for their assistance in this work. And I am specially thankful to my advisor, David Page, for introducing me to the world of sex chromosome biology and evolution, and for many discussions over the years about science and life.

I am also indebted to many collaborators in this work. I would like to thank Jan Hoovers, Jan de Vries and Sjoerd Repping for their good spirit and beautiful FISH analysis of DAZ genes over several years, David Schwartz's laboratory for optical mapping experiments, Christine Disteche and the Whitehead/MIT Genome Center for early collaborations, and Jeanie Maines and Steve Wasserman for enabling us to begin functional studies of DAZ in a model organism.

Finally, I would like to thank Detlev Biniszkiewicz and my family for believing in me and encouraging me through my graduate experience.

Table of Contents

CHAPTER 1

Introduction

Section I. Background

The Y chromosome is one of the smallest of human chromosomes, consisting of a heterochromatic and a euchromatic region. The heterochromatic region is on the distal long arm, and varies drastically in size, from comprising half of the chromosome in some males to being almost absent in others. The euchromatic region consists of about 30 million base pairs, and is responsible for the functions attributed to the Y chromosome. The sex-determining gene, *SRY*, is located on the short arm of the euchromatic Y, and is required for initiating the male sexual pathway of development. At both ends of the Y chromosome lie the pseudoautosomal regions, the regions at which the Y pairs and recombines with the X during meiosis. Abnormalities in pairing between the X and the Y leads to disruption of meiosis and spermatogenic failure. The remaining non-recombining, euchromatic portion of the Y (NRY) has a segmented structure, with regions containing active single copy genes similar to X chromosomal genes, and segments of Y-specific repeats. Y-specific repeats can either be scattered throughout the NRY or clustered in specific regions of the chromosome. X-Y homologous genes are usually ubiquitously expressed, but the Y specific repeats, which often comprise gene families, are testis specific. The major phenotypes associated with the NRY are defects in spermatogenesis, resulting in male infertility. In my thesis work, I examined one of the Y specific repeat families, the *Deleted in Azoospermia* (*DAZ*)gene family, implicated in severe spermatogenic failure and male infertility.

Section II. *Azoospermia factors (AZF)* on the human Y chromosome

The normal male germ cell development pathway in humans must first be appreciated to understand possible causes of spermatogenic failure in men. The indifferent gonad is formed in the embryo by about five weeks of embryonic development, and proliferating primordial germ cells, which are sequestered in extraembryonic tissue during global methylation of the embryo, migrate through the hindgut and populate the gonad around six weeks. By the seventh week, expression of the Y chromosomal gene SRY in somatic tissues of the gonad specifies the male fate, and testis cords, containing spermatogenic precursors encased in tubes with somatic Sertoli

cell precursors, are evident. The spermatogenic precursors, called gonocytes, differentiate and enter a mitotically inactive stage until puberty (Moore and Persaud 1998).

At puberty, spermatogenesis begins in the male, and the process of germ cell division and differentiation into sperm continues throughout life. The entire process takes 65 days and has been extensively studied from a histological perspective. The adult testis has germ cells at various developmental stages within the seminiferous tubules, organized with the earliest germ cell stages at the periphery, and progressively more differentiated germ cells toward the central lumen. The earliest germ cells, the spermatogonial stem cells, are capable of self-renewal and of giving rise to more differentiated, mitotically active spermatogonia that are well characterized morphologically. The most mature spermatogonia undergo DNA replication and differentiate into primary spermatocytes, which then undergo meiosis. During the extended meiotic prophase, genetic recombination between homologous chromosomes occurs, and these stages are well defined according to the state of chromosome condensation. Then spermatocytes undergo the first reductional meiotic division to generate secondary spermatocytes. These enter meiosis II which results in haploid round spermatids. Round spermatids undergo a drastic morphological transformation into elongating spermatids, and these become spermatozoa, which are released into the lumen of the tubule.

A window to identify genes important in human male germ cell development came from observations of Tiepolo and Zuffardi (1976). They observed large Y chromosomal long arm deletions in infertile men and postulated that a factor controlling spermatogenesis was localized to this region. Tiepolo and Zuffardi found that six patients with azoospermia (no sperm in semen) had deletions in Yq. Fathers were tested for the same deletions in four cases, and all fathers had intact Y chromosomes. Therefore, these deletions were *de novo* mutations that caused infertility in the patients. These infertile men were otherwise completely healthy, suggesting that this factor was only required for male fertility. However, these men had no detectable heterochromatin too and could also have had terminal deletions of the Y chromosome. So, the contribution of the Y chromosome to spermatogenesis could have been structural, as failure of the Y to pair at the Yq pseudoautosomal region with the X also results in meiotic arrest and spermatogenic failure. Tiepolo and Zuffardi did not make the distinction between the two hypotheses.

This proposed spermatogenic factor on Yq was later termed the azoospermia factor (*AZF*) as no sperm were observed in semen of men with Y chromosomal long arm deletions (Ma et al. 1992). Further investigation into the *AZF* hypothesis came after many years of work identifying Y chromosomal probes and extensive mapping of the Y chromosome in the late 1980s (Affara et al. 1986). Naturally occurring cytogenetic deletions and structural abnormalities of the Y chromosome were used to order Y chromosomal markers and make a map of the Y chromosome (Vergnaud et al. 1986). By 1992, the Y chromosome was divided into 43 deletion intervals (Vollrath et al. 1992). In addition, a physical map of the entire Y chromosome were published (Foote et al. 1992). Y-specific probes for Southern analysis and PCR-STSs enabled the detection of much smaller deletions in infertile men with cytogenetically normal Y chromosomes (Chandley and Cooke 1994; Ma et al. 1993; Ma et al. 1992).

Multiple AZF regions on the Y chromosome with different Testicular Phenotypes

More sophisticated mapping techniques revealed that there are at least three regions on the Y chromosome required for male fertility (Kobayashi et al. 1994; Vogt et al. 1996) (Figure 1). These three non-overlapping regions of Yq could be independently deleted in infertile men, and were named *AZFa*, *b* and c. (Vogt et al. 1996). Vogt et al reported that testicular phenotype of *AZFa* deleted men was most severe, with a complete absence of germ cells. *AZFb* deletion phenotype was azoospermia with germ cells arrested at meiosis. The *AZFc* deletion testicular phenotype varied widely as discussed below. *AZFc* deletions have been found in 13% of men with azoospermia, whereas deletions of *AZFa* and *AZFb* occur in less than 1% of azoospermic men (Reijo et al. 1995; Vogt et al. 1996). *AZFc* deletions are also observed in 7% of men with severe oligospermia, with sperm counts less than 2 million/ml in semen (Silber et al. 1998). The presence of multiple spermatogenic factors on the Y coincides well with the identification of several testis-specific genes on the Y chromosome. On this basis, Lahn and Page described the Y as a functionally coherent chromosome with roles in male fertility (Lahn and Page 1997).

AZFc deletion phenotype

A range of phenotypes has been observed in males with deletion of the *AZFc* region (Reijo et al. 1995; Vogt et al. 1996). Some *AZFc*-deleted men have no germ cells in seminiferous tubules of their testes (called Sertoli cell only syndrome), whereas other men display testicular maturation arrest, wherein spermatogonial stem cells and pre-meiotic

spermatocytes are detected but spermatogenesis does not proceed beyond meiosis. In a subset of men with arrest in spermatogenesis, however, occasional sperm are observed, indicating that a very small number of stem cells actually do progress through all stages of spermatogenesis to produce mature sperm. *AZFc* deletions have also been detected in several oligospermic men, who have markedly reduced sperm counts (Reijo et al. 1996a). Furthermore, a high degree of phenotypic variation is seen even in different tubules of some men with an *AZFc* region deletion; some tubules have no germ cells whereas others have spermatogenic precursors at intermediate stages of development. The phenotype observed in men with *AZFc* region deletions is restricted to spermatogenesis and these men are usually otherwise healthy, suggesting that any genes involved must either be only expressed in spermatogenesis or be functionally required only for spermatogenesis (Reijo et al. 1995).

Soon after *AZF* region micro-deletions were described, Nagafuchi et al reported an inherited *AZF* deletion, suggesting that some *AZF*-deleted men could produce enough normal sperm for natural fertilization (Nagafuchi et al. 1993) . Subsequently two more cases were described in which men deleted for the *AZFc* region fathered sons carrying the same deletion (Stuppia et al. 1996a, Vogt et al. 1996). A concern has arisen that assisted reproduction techniques propagate the genetic causes of infertility, as any sperm retrieved for intra-cytoplasmic sperm injection from *AZFc*-deleted men may result in *AZFc*-deleted boys who would then become infertile (Reijo et al. 1996a, Silber et al. 1998). Recently, Chang et al reported a remarkable family in which an *AZFc*-deleted man fathered eight children, but whose four sons were all azoospermic and infertile at an early age (Chang et al. 1999). The *AZFc*-deleted father was himself azoospermic when tested at a later age. The ability of an *AZFc*-deleted man to father so many children may be related to his wife's fertility, but this case strongly points to an observation that *AZFc* region deletion phenotypes get progressively worse with age. The variable expressivity of *AZFc* deletions may also be a result of micro-environmental influences in the testes of each man, as often there is variation in phenotype between tubules of the same testis.

This observed range in *AZFc* deletion phenotypes could have several explanations. A more severe phenotype is observed with larger *AZFc* region deletions, suggesting that multiple spermatogenesis genes are responsible for the phenotype. In 1998, Silber et al reported a study of 19 men with Y deletions encompassing the *AZFc* region. They found that all five men with Y

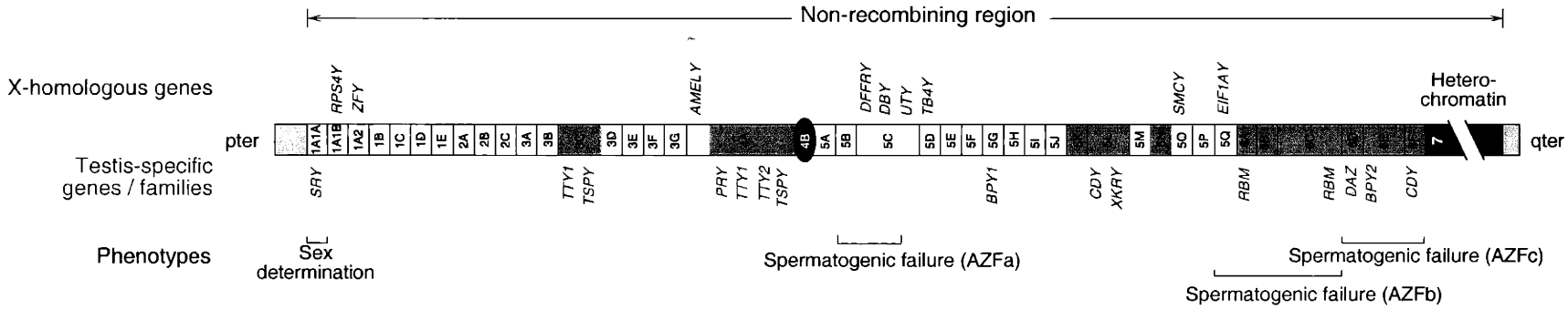


Figure 1. Three regions on the Y chromosome are associated with spermatogenic failure -*AZFa*, *AZFb* and *AZFc*. The Y chromosome is divided into deletion intervals (1-7) based on naturally occurring deletions. Pseudoautosomal regions are indicated in yellow. All known genes in the non-recombining part of the Y chromosome are shown; regions of Y specific repeats are shown in blue. Centromere and heterochromatin on distal Yq are shown in black. (Figure courtesy of B. Lahn)

deletions larger than the *AZFc* region (intervals 6D-6F) had absolutely no sperm in testis biopsies, whereas of 14 men with deletions limited to intervals 6D-6F, all but two (86%) had sperm in testis biopsies (1998). All genes in this region are present in multiple copies, and deletions altering

dosage of different genes in *AZFc* may result in different testicular phenotypes. Breakpoints of *AZFc*-deleted men appear to cluster based on STS analysis, but this technique does not have the power to discriminate between different gene copies. Other factors may affect the expressivity of these deletions, such as local micro-environmental or stochastic processes within the tubules of the testes, modifiers in the genetic background of different men, age of men or environmental factors.

A large number of men with Y deletions have been reported in the literature (Reijo et al. 1995, Najmabadi et al. 1996, Qureshi et al. 1996, Reijo et al. 1996a, Stuppia et al. 1996b, Foresta et al. 1997, Kleiman et al. 1999). Although some of the assays are unreliable and have led to the reporting of single STS 'deletions' which are actually failures of PCR amplification, overall patterns of Y deletions have remained constant. A comprehensive analysis of all available literature on Y chromosomal deletions in infertile men was recently conducted by Laura Brown (pers. comm.). Most frequent deletions reported were of the *AZFc* (51% of all reported deletions in infertile men) or *AZFc/b* region (26%), followed by deletions of the *AZFb* region (8.5%) and then the *AZFa* region (5.6%). The frequency of *AZFc* deletions in azoospermic men remains 12-13% (as first reported by Reijo et al in 1995).

Section III. Gene Content of the *AZF* regions on the Y chromosome

Spermatogenesis genes in the AZFa region

The *AZFa* region of the Y chromosome was recently sequenced and found to span approximately 0.8 Mb (Sun et al. 1999). Two single copy XY homologous genes USP9Y and DBY were found to lie entirely within this deleted region. Mutation screening of 576 infertile men revealed a point mutation in USP9Y that truncates the protein by 90%. This was the first demonstration of an *AZF* phenotype being attributed to a single gene. The azoospermic man with a de novo point mutation of USP9Y had pre-meiotic and meiotic germ cells in most seminiferous tubules, and some post-meiotic sperm in some tubules. This testicular phenotype is

hypospermatogenesis and spermatogenic arrest. Men deleted for the entire *AZFa* region, however, typically have a more severe testicular phenotype, with no germ cells at all (Sertoli cell-only syndrome). It was proposed that loss of DBY accounts for the more severe phenotype of men deleted for both USP9Y and DBY (Sun et al. 1999). Both genes within this region appear to contribute to spermatogenesis in distinct ways. This raises the possibility that in other *AZF* regions too, multiple genes, involved in different aspects of germ cell development, may be responsible for deletion phenotypes observed. *AZFa* was easier to examine in its entirety than *AZFb* or *AZFc*, as it consists only of single copy sequences. Both genes USP9Y and DBY are ubiquitously expressed Y homologues of X-Y common gene pairs, and it will be interesting to see how loss of the Y linked genes results solely in spermatogenic defects.

Candidate Spermatogenesis genes in the AZFb region

The first *AZF* candidate isolated was *RBMY*, a testis-specific transcript that hybridized to a Yq cosmid with a CpG island (Ma et al. 1993). Since then, further analysis of transcripts and genomic regions of *RBMY* suggests that *RBMY* is a gene family consisting of several subfamilies and over ten members, with sequences on both arms of the Y chromosome (Prosser et al. 1996, Chai et al. 1997). These genes are expressed exclusively in the male germline in both man (Elliott et al. 1997) and the mouse (Elliott et al. 1996). Seven *RBMY1* genes with slightly different sequences have been identified, and they encode proteins with a RRM type putative RNA binding domain and four copies of an internal 37 amino acid repeat termed the SRGY repeat (Chai et al. 1998). The *RBMY2* genes of a second subfamily have 88% homology with *RBMY* and contain only one SRGY repeat. These genes are expressed at a low level. Other subfamilies are thought to be pseudogenes (Chai et al. 1998). Functional *RBMY1* genes appear to be clustered in the *AZFb* region. Men with *AZFb* deletions have no testicular *RBMY1* transcripts detectable by in situ hybridization suggesting that this region contains all functional *RBMY1* genes (Elliott et al. 1997). RBM family proteins are nuclear proteins abundant in primary spermatocytes, the most mature cell-types seen in *AZFb*-deleted men (Vogt et al. 1992).

Candidate spermatogenesis genes within the AZFc region

Deleted in Azoospermia (DAZ) was initially identified as an *AZFc* candidate in a systematic exon trapping search for genes in the *AZFc* region (Reijo et al. 1995). For this gene

hunt, a map of the *AZFc* region was constructed by STS analysis of YAC and cosmid clones, and the repeat rich-nature of the region was acknowledged but not sufficiently understood. *DAZ* was thought to be a single copy gene, based on hybridization patterns on a Southern blot. Northern analysis revealed an abundant 3.5 kb transcript in the testis. The *DAZ* coding region has one or multiple N terminal domains matching the RNP/RRM consensus observed in many proteins that bind RNA or single stranded DNA, and multiple tandem repeats of a 24 aa unit with a high concentration of proline, glutamine and tyrosine residues, as is typical of many RNP/RRM proteins (Burd and Dreyfuss 1994, Kenan et al. 1991). My thesis work has focused on characterizing the structure and evolution of this gene family in depth (Chapters 2, 3, 4) and on examining if *DAZ* has a role in meiosis (Appendix 2). We identified four nearly identical *DAZ* genes in two clusters within the *AZFc* region, and a highly similar *DAZ* homolog (*DAZL*, formerly called *DAZH*) on chromosome 3 (Saxena et al. 1996). *DAZ* and *DAZL* proteins have high amino acid identity, especially in the RNA-binding domains. The amino acid residues forming the RNP1 and RNP2 peptides, which are thought to mediate actual RNA binding, are identical in the two proteins. The major difference between the two proteins is that the RNA-binding domains are multimerized in some *DAZ* proteins, and that on the C-terminal end, *DAZ* proteins have multiple copies of a 24 amino acid repeat, whereas *DAZL* just has one. The *DAZ* genes are expressed exclusively in the testis. *DAZL* is similarly expressed predominantly in testes, but is also expressed at a much lower level in the ovaries. *DAZ* gene family transcripts in adult testes have been localized to spermatogonial stem cells and pre-meiotic spermatocytes (Menke et al. 1997).

Other genes within the *AZFc* region were later identified and shown to be germ cell specific (Lahn and Page 1997). The *CDY1* gene, which landed in this region by retroposition of a processed transcript (Lahn and Page 1999b) encodes a testis-specific histone-acetyl transferase. The *AZFc* region has two copies of this gene, and another active copy is found elsewhere on the Y chromosome (Y chromosome sequencing project, T. Kawaguchi, H. Skaletsky and D.C. Page unpublished observations). Repackaging of sperm chromatin is an essential part of post-meiotic spermiogenesis, and CDY1 protein may aid in removal of histones from DNA to free DNA for binding to protamines (Lahn and Page 1999b). Loss of CDY1 likely contributes to the *AZFc* phenotype, but is not the sole contributor, as CDY is not expressed until after meiosis, and the phenotypes appear to affect earlier testicular cells. The *BPY2* gene, encoding a small testis-

specific basic protein, is present in three copies within the *AZFc* region (Y chromosome sequencing project, H. Skaletsky and D.C. Page unpublished observations). The role of this gene in *AZFc* deletions has not been examined. *RBMY* genes are also found in the *AZFc* region, but whether these are functionally active copies, or pseudogenes, is undetermined (Glaser et al. 1997) (Y chromosome sequencing project, H. Skaletsky and D.C. Page unpublished observations). Other genes may still reside in the *AZFc* region that are as yet unidentified. Gene hunting in the *AZFc* region is still underway, clones covering 560 kb were recently subjected to exon-trapping and 11 potential new exons were identified but it remains to be seen whether any new *AZFc* genes will emerge from this search (Wong et al. 1999).

No mutations or internal deletions limited to *DAZ* or any other *AZFc* region candidate genes in infertile men have thus far been identified. However, these genes are all present in multiple copies within *AZFc*, and for *DAZ*, *CDYI* and *BPY2* all copies appear capable of encoding functional proteins. Therefore, it is possible that disruption of one copy of any gene alone may not result in an infertility phenotype. In Chapter 4, I will describe that fertile men from some populations may have Y chromosomes with two copies of *DAZ* (and, as these men have large deletions, copy numbers of other genes may also be fewer).

AZFc deletions remove multiple testis-specific genes that might all be involved in different aspects of spermatogenesis. Therefore, absence of each gene product may contribute in different ways to the overall spermatogenic defects observed in *AZFc*-deleted men. *RBMY* and *CDY* copies are also present elsewhere on the chromosome, so the contribution to the *AZFc* phenotype is likely to result from lowered protein levels. If *DAZ* and *DAZL* proteins have overlapping functions, the contribution of *DAZ* to the *AZFc* phenotype is also likely a matter of lowered protein level. This view of the *AZF* region deletions as more classic genomic disorders (Lupski 1998), removing multiple genes involved in different aspects of spermatogenesis, is more consistent with the variable *AZF* phenotype than the disruption of one essential spermatogenesis gene. If in *AZFc* deleted patients, levels of these proteins are at the threshold of what is required for different aspects of male germ cell development, then depending on local concentrations of protein levels, spermatogenesis could proceed to different extents in different spermatogenic tubules.

In an attempt to recreate an *AZFc* like phenotype in mice, Vogel et al recently generated a mutant mouse model heterozygous for the *DAZ* homologue *Dazl* (described below) and carrying

a 2-4Mb Y chromosomal deletion (Ydl) encompassing most mouse *Rbm* genes, and perhaps other genes (Vogel et al. 1999). In these mice, a phenotype similar to *AZFc* region deletion in humans was observed, with significantly reduced sperm counts, and variable testicular histology in different individual mice studied. Some mice had all stages of spermatogenesis, but reduction of numbers of elongating spermatids, whereas others had some tubules with only Sertoli cells, and others with pachytene spermatocyte arrest. Although this phenotype could have arisen from the reduction in *Dazl* and other as yet uncharacterized Y chromosomal genes within the Ydl chromosome, this experiment confirms that a lowered dosage of Dazl protein can contribute to an *AZFc* like-phenotype.

Section IV. Structure of the *AZFc* region and mechanism of *AZF* region deletions

Recent mapping and sequencing analysis of the *AZFc* region from our lab suggests that it spans approximately 3.4 Mb, over 10% of the chromosome (Y chromosome sequencing project, T. Kawaguchi, H. Skaletsky and D.C. Page unpublished observations). This region of the chromosome is susceptible to recurring deletions, which either result in infertility (*AZFc* deletions), or are naturally occurring polymorphisms (Jobling et al, 1996, chapter 4). This propensity for deletion (and duplication or inversion) is not surprising as this region is a complicated mosaic of repeated elements each extending over hundreds of kilobases (discussed in Appendix 1).

AZF region deletions are mostly newly arising mutations in the male germ line and are selected out of the population within one or two generations. Two homologous10kb retroviral elements with a 2 kb region of perfect overlap flanking the *AZFa* region are responsible for deletion in at least some *AZFa* deleted males (Sun et al, in preparation, pers. comm.). The regions of homology around the *AZFc* region, and within this commonly deletion are much larger (around 250 kb) and differ even less (Y chromosome sequencing project, H. Skaletsky and D.C. Page unpublished observations). Therefore deletions (and duplications or inversions) may be more likely to occur in these regions. Since the non-recombining part of the Y chromosome does not have a partner in meiosis, *AZF* region deletions likely occur by unequal sister chromatid exchange or excision of an intra-chromosomal loop during mitosis in spermatogonia or during meiosis. The combined estimated mutation rate of the *AZF* loci, which is about 1/8000 men

(1.25×10^{-4}), is comparable to the mutation rate of the X chromosomal DMD and STS loci, both of which have deletion as their major mutation events, and about ten times higher than usually found in autosomal loci (Reijo et al. 1995). The human and mouse Y chromosomes are seen to associate with itself, back-folding, in meiosis, and this may be due to interchanges within the chromosome. (Chandley and Speed 1987; Edwards and Bishop 1997). No hot-spots for *AZFc* deletions have been identified so far. I will describe my effort in mapping AZFc region breakpoints in Appendix 1.

Section V. Theories of Y chromosome degenerative evolution

Degeneration of the Y chromosome - theories and experimental observations

The system of sex determination seen in mammals is thought to have arisen from one member of a pair of autosomes acquiring a gene that was sufficient to trigger the male pathway of sexual differentiation (Bull 1983). That chromosome became the Y, and its partner the X. At first, recombination occurred throughout the length of both sex chromosomes, but ceased in a step-wise fashion, likely because of multiple inversion events on the Y chromosome (Lahn and Page 1999a). The highly differentiated human X and Y chromosomes only recombine at the pseudoautosomal regions at both ends of the chromosome. After cessation of recombination throughout most of the length of the Y chromosome, it is observed that most genes it shared with the X chromosome decay (Rice 1996).

Multiple evolutionary theories have been proposed to explain why Y chromosomes deteriorate. The first proposed theory was Muller's sheltered-lethal model (Muller 1918). Recessive mutations in Y linked genes would never be expressed because the Y always remains heterozygous. Hence, such mutations would be tolerated and lead to decay of Y-linked genes. Fisher (Fisher 1935) refuted this theory by quantitative analysis of Muller's model, suggesting that enough X-linked mutations would be present in a large population to unmask Y recessive alleles and cause them to be selected out of the population. Thus loss of function mutations on the Y could not accumulate.

Population genetic theories used to explain the case for extinction of asexual species are alternative theories now used to explain Y degeneration. Charlesworth adapted a theory proposed by Muller, termed Muller's ratchet, to the Y chromosome in 1978 (Charlesworth 1978). The theory states that without recombination, errors would accumulate continuously by chance (whereas by recombination, intact alleles could be regenerated by crossovers between mutated alleles), and the only means of correcting errors would be low-probability reversions. Random loss of the chromosome with no or few mutations would begin a ratchet; more and more mutations would arise on an already mutated chromosome. Rice tested this theory of Y degeneration experimentally, by creating a large proto-Y in a small *Drosophila* population, and examining this chromosome after 35 generations. He found that the Y chromosome had degenerated, and was also physically smaller. Processes in addition to Muller's ratchet may have caused this decay, but it was likely the major contributor (Rice 1994). A second theory, adapted to the Y chromosome by Rice (Rice 1987), states that without recombination, the spread of a selectively advantageous Y allele would necessitate the fixation of deleterious alleles that happened to be on other loci linked to the chromosome on which the favorable allele originated. This theory is termed genetic hitchhiking, as the allele that enhances fitness is forced to bring along the rest of the chromosome, the "hitchhiker".

Accumulation of Male Fitness Genes on the Y chromosome

A hypothesis put forward by Fisher (1931) to explain the evolution of suppressed recombination on the Y chromosome was based on the accumulation of sexually antagonistic alleles at loci tightly linked to the sex-determining locus. Although Fisher proposed this hypothesis for an alternative purpose, his hypothesis has, until the work presented in Chapter 2, been the only consideration of male benefit alleles accumulating on the Y chromosome. Sexually antagonistic alleles are loci with variants favoring one sex, but are detrimental to the other sex. Fisher proposed that the constraints for accumulation of sexually antagonistic mutations are fewer surrounding the sex-determining locus, as there is sex-biased transmission of this locus. Suppression of recombination would be one way to prevent expression of sexually antagonistic alleles in females (Fisher 1931).

Fisher based his hypothesis on genetic mapping studies of male ornamentation traits of the guppies performed by Winge in 1927. Winge mapped 18 major genes producing ornamental

characteristics such as bright color, long tails, spots of varying color and size, and found, that remarkably,17 of these 18 genes mapped to the sex chromosomes, very close to the sex-determining locus. As guppies have 23 pairs of chromosomes, this clustering of ornamentation genes was attributed to 'hot spots' by Winge (Fisher 1931). Fisher assumed that the ornamentation would make fish more conspicuous to predators in both sexes, but is advantageous to males for mating, so would tend to accumulate on the Y chromosome.

Section VI. The Y chromosome and studies of human evolution

The non-recombining portion of the Y chromosome and mitochondrial DNA provide simple records of the paternal and maternal origins of people. Neither of these systems recombines during meiosis, so they carry a trace of the ancestor from which they originated. The Y chromosome has been used extensively for human evolution and population studies over the past few years (Hammer 1995) It has been a powerful tool for tracing and comparing paternal lineages of human populations, based on the identification of polymorphic markers evolving at different rates. Y chromosomal analysis has been applied to answer questions on the origin of modern humans and their spread to different continents (Hammer et al. 1998), the contribution of men with diverse Y chromosomes to specific population admixtures (Zerjal et al. 1997), to identify genetic affiliations of linguistically related populations, and to study lineages descended from one man.

However, there appears to be a scarcity of polymorphisms on the Y chromosome, as systematic searches have revealed far fewer variations than expected (Jobling and Tyler-Smith 1995; Poloni et al. 1997b). This finding is contradictory to expectations because, as the Y is gene poor, it should accumulate mutations and junk transposable elements more readily than other regions of the genome. Two alternative theories have been put forward as explanations: i) that the effective Y population size is small, with some males having many offspring and most others none or few, thus restricting diversity on the Y, or ii) that selective sweeps have occurred in the history of the Y, where selectively advantageous mutations arising in a Y specific gene spread through the population and the rest of the Y chromosome with no variation is brought along as a 'hitchhiker' (Jobling and Tyler-Smith 1995).

The initial discoveries of polymorphic markers on the Y chromosome came slowly, and few RFLP markers were found (Casanova et al. 1985; Lucotte et al. 1991), but in recent years, there has been much more nucleotide substitution and microsatellite repeat variation detected on the Y chromosome (Jobling and Tyler-Smith 1995, Underhill et al., 1997). NRY polymorphic markers now range from base substitutions and insertion/deletion polymorphisms, which are biallelic in nature and represent rare or unique events in evolution, to microsatellite and minisatellite markers, which are subject to faster mutation rates.

One of the first and most polymorphic markers identified on the Y chromosome was the 49f/a marker, which detected a complex pattern of over 15 bands on *Taq*I southern blots of male DNAs (Lucotte and Ngo 1985; Ngo et al. 1986). Of these, eight bands were found to be polymorphic, and 4000 Y chromosomes from over 70 different human populations were typed for these bands. 49f/a haplotyping was used for studies of West African, Southern African, and central African populations, for multiple studies on several Jewish communities, and for genetic affinity of Australian aboriginal populations, Indian populations and Polynesian populations. A correlation was found between p49f/a TaqI polymorphism haplotypes and linguistic groups of populations (Poloni et al. 1997a). In 1996, we reported that the DYS1 locus detected by p49f/a was *DAZ* (Chapter 2).

Although p49f/a was used extensively early when few Y chromosomal markers were available, the haplotypes observed were difficult to interpret because the molecular basis of the polymorphism was too complex to understand (Jobling and Tyler-Smith 1995). Polymorphic bands could be absent or new bands could arise, and there was a large range of size variation of bands. Furthermore, apparently identical bands could arise independently. Groups who worked with p49f haplotypes could not identify the mutational events giving rise to different alleles, and thus assumed single events for each allelic variation. Analysis relied mostly on estimation of allele frequencies in populations (Poloni et al. 1997a). In chapter 4, I will describe the molecular basis of most polymorphic 49f/a *Taq*I fragments and estimate the rate of change in *DAZ* genes in human populations using 49f polymorphism analysis.

Section VII. *Deleted in Azoospermia* homologues in model organisms

Since the identification of *DAZ* in humans, homologues of *DAZ* and *DAZL* have been identified in a variety of model organisms and found to be expressed only in the germ line of all organisms studied. Y-linked *DAZ* homologues are present only in Old World monkeys, as *DAZ* arose in the primate lineage by transposition of an ancestral autosomal gene (Chapter 2), (Saxena et al. 1996). Genetic analysis has revealed that *DAZ* homologues in several vertebrate and invertebrate systems are essential for different aspects of germ cell development in one or both sexes, discussed below.

The autosomal mouse *DAZL* locus was identified by hybridizing human *DAZ* probes, and found to be transcribed in male and female germ cells as early as E11.5, before the onset of sex differentiation (Cooke et al. 1996; Reijo et al. 1996b). Levels of expression in the male gonad increase as embryonic and post-natal development proceeds and plateau at day 10 after birth, when spermatocytes enter meiosis (Reijo et al. 1996b). In situ hybridization reveals that the highest levels of *DAZL* transcripts in adults are in spermatogonia about to divide and enter meiosis (Niederberger et al. 1997). The *DAZL* protein is most abundant in pachytene spermatocytes (Ruggiu et al. 1997; Reijo et al, submitted) although it is also detectable in late type B spermatogonia, pre-leptotene spermatocytes and zygotene spermatocytes. Vogt et al have also reported protein expression in sperm tails (Habermann et al. 1998).

Targeted disruption of the autosomal *DAZL* locus results in male and female infertility (Ruggiu et al. 1997). The earliest phenotype observed is embryonic loss of germ cells starting around E15, when cell divisions have ceased in germ cells of both sexes. At this stage, female germ cells have arrested in prophase of meiosis I, and male germ cells (prospermatogonia) are in a quiescent phase after undergoing proliferative mitotic divisions. In adult mutant males, testis weights are about one third that of normal males. Testicular histology of *DAZL*-deficient adults reveals no germ cells beyond the spermatogonial stage and fewer tubules per cross section than in wild-type or heterozygous males. Meiotic germ cells, the spermatocytes are completely absent. Heterozygous *DAZL* mutants are fertile but have high numbers of abnormal sperm and sub-normal (2/3 of normal) testis weights (Ruggiu et al. 1997).

The *Xenopus DAZL* homologue *xDAZL* was identified as a component of *Xenopus* germ plasm (Houston et al. 1998).*xDAZL* RNA is expressed in the mitochondrial cloud of stage I

oocytes and maternal *xDAZL* mRNA remains expressed in embryonic germ plasm until neurulation. xDAZL protein also localizes to germ plasm of embryos, becoming detectable in blastula stages and persisting until early tailbud stages. xDAZL protein is cytoplasmic and appears to be under translational repression. It becomes undetectable during migration of primordial germ cells (Houston and King 2000). In adults, *xDAZL* RNA is abundantly expressed in spermatogonia and spermatocytes of the testis, and in ovaries, but not in any somatic tissues. In vitro, xDAZL protein can bind homopolymeric stretches of polyG and polyU RNA but not polyA and polyC RNA (Houston and King 2000)

An essential role for xDAZL in PGC (primordial germ cell) development was recently demonstrated by Houston and King (2000), who specifically depleted maternal *xDAZL* RNA in oocytes using anti-sense oligonucleotides. Tadpoles from oocytes depleted of *xDAZL* have few or no migrating PGCs, and gonads from later stage tadpoles have no germ cells. All effects were shown to be specific to *xDAZL* depletion, as injection of *xDAZL* RNA rescues PGC development. Analysis of germ cells using other markers show that *xDAZL* depleted PGCs fail to migrate to the dorsal endoderm at correct time and are lost. Houston and King proposed that xDAZL is required during PGC differentiation for establishing migrational competence of an early population of PGCs. Since xDAZL is likely to be a RNA binding protein, xDAZL could regulate translation or expression of proteins needed to respond to migrational cues, or repress expression of factors inhibiting migration.

The nematode *C. elegans* is the most distantly related organism to man in which a *DAZL* homologue has been studied. *C. elegans daz*-1 was identified from the *C. elegans* genome and cDNA projects based on homology to mammalian DAZ proteins. Experiments from M Yamamoto's lab in Japan have demonstrated an essential role for this protein in oogenesis (A. Sugimoto, pers. comm.). daz-1 protein is expressed in hermaphrodite adult gonads starting at the mitosis/meiosis transition zone and increasing in intensity until the pachytene region. RNA mediated interference experiments showed that F1 hermaphrodite progeny were sterile at high penetrance, but their sibling males were fertile and had normal gonads and sperm. Gonads of F1 hermaphrodites had normal-looking sperm, but oogenesis was affected, with most cells in the gonad appearing mitotic and an absence of meiotic cells. Occasional pachytene nuclei were observed, but no later stages of meiosis were evident. This phenotype was confirmed by identification of a *daz*-1 mutant by screening a library of mutants generated by transposon

insertion. The authors concluded that *daz*-1 seemed to function during oogenesis in either the transition from mitosis to meiosis, or in the early stage of meiosis.

Zebrafish *DAZL* (*zDAZL*) was identified by homology to other *DAZ* genes, and its expression characterized by Maegawa et al (1999). zDAZL is a 229 aa protein. *zDAZL* RNA is strongly expressed in both ovary and testis and not in any other tissues. Expression analysis in the embryo revealed that maternal *zDAZL* mRNA localizes to the vegetal pole of the 1 cell embryo, and migrates through cytoplasmic streams to blastomeres as early embryogenesis proceeds, and disappears in late blastula stage (as in *Xenopus* embryos). In situ hybridization of ovary sections shows localization of *zDAZL* mRNA to the cortex of the oocyte, suggesting that *zDAZL* is restricted during oogenesis to a region which will become the vegetal pole of the embryo (Maegawa et al. 1999).

The *Drosophila DAZL* homologue, *boule*, was identified independently from *DAZ*, in a screen for transposon induced male sterile mutants (Eberhart et al. 1996). The boule protein has 42% amino acid identity to human *DAZ* and *DAZL*, and contains a single RNP-type RNA binding domain. Two mutant alleles were identified, and both were found to be nulls. Northern analysis showed that in adults, *boule* is expressed only in the male testis. *Boule* expression during development, starting during larval stages and continuing till adulthood, coincides with the onset of testis development and spermatogenesis in males. *Boule* homozygous males, although sterile, have no other visible defects, and *boule* mutant females are fertile . The *boule* mutant phenotype will be discussed in detail in following sub-section.

Spermatogenesis in Drosophila and the boule male sterile mutation

An overview of the process of spermatogenesis in flies is now given, as this model system was used to assay the meiotic role of mammalian *DAZ* proteins (Appendix 2). In *Drosophila melanogaster* spermatogenesis begins in the germinal proliferation center at the apical tip of the testis. Germ line stem cells are enclosed in a pair of cyst progenitor cells and form a rosette around apical tip cells that lie at the conical tip of the testis. Spermatogenic differentiation starts when one germline stem cell attached to hub divides asynchronously to give rise to a daughter stem cell and a primary spermatogonial cell. The cyst progenitor cells also divide at this time, and the newly derived primary spermatogonial cell becomes associated with a pair of cyst cells that enclose the progeny of that cell throughout spermatogenesis. The cyst,

composed of two cyst cells and the germ cells enclosed within them, proceeds through spermatogenic differentiation as a unit, with nearly complete synchrony. Within the cyst, the spermatogonial cell undergoes four mitotic divisions to result in sixteen primary spermatocytes. Premeiotic DNA replication happens early after the last mitotic division, and primary spermatocytes then undergo a prolonged G2 phase, lasting over 90 hours, of cell growth and gene expression. The volume of spermatocytes increases 25-fold. No recombination or synaptonemal complex formation occurs in *Drosophila* males so classical meiotic stages such as the leptotene, zygotene and pachytene are absent. However, during this growth period genes required for meiosis and spermatid differentiation are transcribed. The first and second meiotic divisions occur in rapid succession after completion of the growth phase. In MI, the reductional division, the XY pair and three autosomes segregate, and MII is quite similar to the mitotic divisions. Cytokinesis is incomplete throughout spermatogenesis so that all 16 mitotic and meiotic daughter cells in a cyst are connected throughout development. After completion of meiosis, the 64 haploid germ cells carry out an elaborate differentiation program. The nucleus reforms in each spermatid, and mitochondria fuse to form a mitochondrial derivative, the nebenkern. As each spermatid bundle migrates to the base of the testis, the nuclei become reduced in volume and change from spheres to long thin cylinders. The sperm tail is generated, and flanked by the elongating nebenkern. Sperm individualization and coiling follow, and mature sperm exit the basal end of the testis into the seminal vesicle for storage (Fuller 1993).

The study of meiotic entry and progression in *Drosophila* males is based on detailed morphological analysis of germ cells before, during and after the two meiotic divisions. Cysts undergoing meiosis are located at about one third of the length from the apical tip of the testis. Late primary spermatocytes at the end of growth phase have a large nucleus with a prominent nucleolus. The XY pair and the second and third bivalent are seen as distinct structures at the periphery of the nucleus. In early prometaphase, the nucleus rounds up, the nucleolus becomes smaller and paler and an aster-like array appears at one side of the nucleus. The chromosomes condense, resulting in one round, and one oblong pair of autosomes. The nucleolus then breaks down as the nucleus becomes irregularly shaped. The centrosome divides into two, and moves to form the two poles of the meiotic spindle. Chromosomes move from the nuclear periphery to the metaphase plate. Cyclin A moves from the cytoplasm to the nucleus. At metaphase, the nuclear lamins disperse, and the distinction between nucleus and cytoplasm is blurred. The

completely condensed bivalent chromosomes become attached to spindle fibers attached to asters from opposite sides of the nucleus, cyclin A gets degraded and homologous chromosome segregation occurs. Chromosomes move rapidly to opposite poles in anaphase of meiosis I, and the central spindle composed of inter-digitated microtubules, becomes prominent in late anaphase. At telophase, the spindle has an hourglass shape, and cytokinesis occurs (Cenci et al. 1994).

The *Drosophila boule* mutant has a male sterile phenotype, with defects in the regulation of entry into meiosis (Eberhart et al. 1996). Testicular histology of *boule* mutant males reveals consistent meiotic arrest. Spermatocytes have a wild-type morphology, and premeiotic S phase occurs normally. The growth phase of spermatocytes appears normal by morphological analysis. Meiotic prophase appears wild-type in *boule* mutants, but subsequent stages are disrupted, resulting in cysts containing 16 tetraploid cells instead of 64 haploid spermatids. Chromosomes begin to condense, and cyclin A enters the nuclei, signaling the transition between the G2 and M stages. Cyclin A normally degrades rapidly after translocation into the nucleus but remains detectable in *boule* nuclei. Centrosome separation occurs, but centrosomes do not reach the poles or form asters. The nuclear lamina fails to break down, and nuclear lamins appear intact. These observations suggest that the transition from prophase to metaphase of Meiosis I does not occur in *boule* mutants. Spermiogenesis, the differentiation of spermatids, does continue to an extent in *boule* mutants, and tetraploid spermatids begin to form onion-like nebenkern structures.

To show that *boule* is active in the germ line, Eberhart et al expressed the *boule* protein using a testis specific expression system (Eberhart et al. 1996). The *boule* coding sequence was cloned into a P element vector containing sequences from the b2-tubulin locus that drive expression in postmitotic germ cells of the testis (Hoyle et al. 1995). In *boule* mutant flies, this construct resulted in meiotic products in more than half of the spermatid cysts, and all cysts had extensive spermiogenesis. This partial rescue suggested a dosage sensitive role for *boule* in male germ cell development.

Boule mutants show a phenotype similar to two other *Drosophila* mutants, *pelota* and *twine*, although spermatid differentiation proceeds to more advanced stages in these two mutants as compared to *boule* (Maines and Wasserman, 1998). *Twine* encodes the meiosis specific homologue of Cdc25 phosphatase, and controls the transition from the extended G2 phase into the first meiotic division . cdc25 phosphatases are important mitotic cell cycle regulators

conserved from yeast to man, and in the budding yeast activate the p34cdc2-cyclin kinase complex by removing inhibitory phosphates from p34cdc2 (Coleman and Dunphy 1994). *Pelota* encodes a protein similar to Dom34, and belongs to a family of translation factors (Eberhart and Wasserman 1995). Recently, Maines and Wasserman showed that *boule* is required for proper translation of twine mRNA, and heterologous expression of Twine protein rescues the *boule* meiotic entry defect (Maines and Wasserman 1999).

Potential roles of DAZ/DAZL in germ cell development

The *DAZ* gene family plays multiple roles in germ cell development in different organisms, being essential for early germ cell processes in both males and females in frogs and mice, and required for meiosis in opposite sexes in *Drosophila* and *C. elegans*. It is likely that *DAZ* family proteins have a fundamental role in germ cell development, and took on additional functions in more complex organisms. Therefore, it would not be surprising if *DAZ* proteins in humans have roles in early germ cell development (as seen in mice), during meiosis (as seen in flies), and in post-meiotic spermiogenesis (as indicated in mice heterozygous for *DAZL*, and in flies). It is possible that there may be a partitioning of functions between *DAZ* and *DAZL* in this respect. Several other RNP family proteins are important in the regulation of spermatogenesis in *Drosophila* and in vertebrates, and these are generally required during postmeiotic differentiation, when there is substantial translation of stored mRNAs (Hecht 1993, Hecht 1998, Karsch-Mizrachi and Haynes 1993). Translational repression of protamines, and other RNAs required for spermiogenesis and testis specific splicing of transcripts appear common during germ cell development, so it is likely that RNA binding proteins are important components of the germ cell machinery.

References

Affara NA, Florentin L, Morrison N, Kwok K, Mitchell M, Cook A, Jamieson D, et al (1986) Regional assignment of Y-linked DNA probes by deletion mapping and their homology with X-chromosome and autosomal sequences. *Nucleic Acids Research* **14**:5353-5373.

Bull JJ (1983) Evolution of Sex Determining Mechanisms. In: Wake DB, Slatkin MW (eds) Evolution. Benjamin/Cummings, Menlo Park, CA

Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**:615-621.

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, et al (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* **230**:1403-6.

Cenci G, Bonaccorsi S, Pisano C, Verni F, Gatti M (1994) Chromatin and microtubule organization during premeiotic, meiotic and early postmeiotic stages of *Drosophila melanogaster* spermatogenesis. *Journal of Cell Science* **107**:3521-34.

Chai NN, Salido EC, Yen PH (1997) Multiple functional copies of the RBM gene family, a spermatogenesis candidate on the human Y chromosome. *Genomics* **45**:355-361.

Chai NN, Zhou H, Hernandez J, Najmabadi H, Bhasin S, Yen PH (1998) Structure and organization of the RBMY genes on the human Y chromosome: Transposition and amplification of an ancestral autosomal hnRNPG gene. *Genomics* **49**:283-289.

Chandley AC, Cooke HJ (1994) Human male fertility--Y-linked genes and spermatogenesis. *Human Molecular Genetics* **3**:1449-1452.

Chandley AC, Speed RM (1987) Cytological evidence that the Sxr fragment of XY,Sxr mice pairs homologously at meiotic prophase with the proximal testis-determining region. *Chromosoma* **95**:345-9.

Chang PL, Sauer MV, Brown S (1999) Y chromosome microdeletion in a father and his four infertile sons. *Human Reproduction* **14**:2689-94.

Charlesworth B (1978) Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences* **75**:5618-5622.

Coleman TR, Dunphy WG (1994) Cdc2 regulatory factors. *Current Opinion in Cell Biology* **Dec;6**:877-82.

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. *Human Molecular Genetics* **5**:513-516.

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

Eberhart CG, Wasserman SA (1995) The pelota locus encodes a protein required for meiotic cell division: an analysis of G2/M arrest in Drosophila spermatogenesis. *Development* **121**:3477-86.

Edwards RG, Bishop CE (1997) On the origin and frequency of Y chromosome deletions responsible for severe male infertility. *Molecular Human Reproduction* **3**:549-554.

Elliott DJ, Ma K, Kerr SM, Thakrar R, Speed R, Chandley AC, Cooke H (1996) An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. *Human Molecular Genetics* **5**:869-874.

Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, et al (1997) Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proceedings of the National Academy of Sciences* **94**:3848-3853.

Fisher RA (1931) The Evolution of Dominance. *Biological Review* **6**:345-368.

Fisher RA (1935) The sheltering of lethals. *American Naturalist* **69**:446-455.

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

Foresta C, Ferlin A, Garolla A, Rossato M, Barbaux S, De Bortoli A (1997) Y-chromosome deletions in idiopathic severe testiculopathies. *Journal of Clinical Endocrinology and Metabolism* **82**:1075-1080.

Fuller M (1993) Spermatogenesis. In: Bates AM-AaM (ed) Development of Drosophila. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 61-147

Glaser B, Hierl T, Taylor K, Schiebel K, Zeitler S, Papadopoullos K, Rappold G, et al (1997) High-resolution fluorescence in situ hybridization of human Y-linked genes on released chromatin. *Chromosome Research* **5**:23-30.

Habermann B, Mi H-F, Edelmann A, Bohring C, Bäckert I-T, Kiesewetter F, Aumüller G, et al (1998) *DAZ* (Deleted in AZoospermia) genes encode proteins located in human late spermatids and in sperm tails. *Human Reproduction* **13**:363-369.

Hammer MF (1995) A recent common ancestry for human Y chromosomes. *Nature* **378**:376-378.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, et al (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* **15**:427-41.

Hecht NB (1993) Gene expression during male germ cell development. In: Desjardins C, Ewing LL (eds) Cell and Molecular Biology of the Testis. Oxford University Press, New York, pp 400-432

Hecht NB (1998) Molecular mechanisms of male germ cell differentiation. *Bioessays* **20**:555-61.

Houston DW, King ML (2000) A critical role for Xdazl, a germ plasm-localized RNA, in the differentiation of primordial germ cells in Xenopus. *Development* **127**:447-456.

Houston DW, Zhang J, Maines JZ, Wasserman SA, King ML (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.

Hoyle HD, Hutchens JA, Turner FR, Raff EC (1995) Regulation of beta-tubulin function and expression in Drosophila spermatogenesis. *Developmental Genetics* **16**:148-170.

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet* **11**:449-56.

Karsch-Mizrachi I, Haynes SR (1993) The Rb97D gene encodes a potential RNA-binding protein required for spermatogenesis in *Drosophila*. *Nucleic Acids Research* **21**:2229-2235.

Kenan DJ, Query CC, Keene JD (1991) RNA recognition: towards identifying determinants of specificity. *Trends in Biochemistry* **16**:214-220.

Kleiman SE, Yogev L, Gamzu R, Hauser R, Botchan A, Lessing JB, Paz G, et al (1999) Genetic evaluation of infertile men. *Human Reproduction* **14**:33-38.

Kobayashi K, Mizuno K, Hida A, Komaki R, Tomita K, Matsushita I, Namiki M, et al (1994) PCR analysis of the Y chromosome long arm in azoospermic patients: evidence for a second locus required for spermatogenesis. *Human Molecular Genetics* **3**:1965-1967.

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

Lahn BT, Page DC (1999a) Four evolutionary strata on the human X chromosome. *Science* **in press**

Lahn BT, Page DC (1999b) Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nature Genetics* **21**:4290433.

Lucotte G, David F, Mariotti M (1991) Nucleotide sequence of p49a, a genomic Y-specific probe with potential utilization in sex determination. *Molecular and Cellular Probes* **5**:359-363.

Lucotte G, Ngo NY (1985) p49f, A highly polymorphic probe, that detects *Taq*1 RFLPs on the human Y chromosome. *Nucleic Acids Research* **13**:8285.

Lupski J (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* **14**: p417-22.

Ma K, Inglis JD, Sharkey A, Bickmore WA, Hill RE, Prosser EJ, Speed RM, et al (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, Sharkey A, Kirsch S, Vogt P, Keil R, Hargreave TB, McBeath S, et al (1992) Towards the molecular localisation of the AZF locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. *Human Molecular Genetics* **1**:29-33.

Maegawa S, Yasuda K, Inoue K (1999) Maternal mRNA localization of zebrafish DAZ-like gene. *Mechanisms of Development* **81**:223-226.

Maines JZ, Wasserman SA (1999) Post-transcriptional regulation of the meiotic Cdc25 protein Twine by the Dazl orthologue Boule. *Nature Cell Biology* **1**:171-174.

Menke DB, Mutter GL, Page DC (1997) Expression of *DAZ*, an azoospermia factor candidate, in human spermatogonia. *American Journal of Human Genetics* **60**:237-241.

Moore KL, Persaud TVN (1998) The Developing Human - Clinically Oriented Embryology. W. B. Saunders Co., Philadelphia

Muller HJ (1918) Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics* **3**:422-499.

Nagafuchi S, Namiki M, Nakahori Y, Kondoh N, Okuyama A, Nakagome Y (1993) A minute deletion of the Y chromosome in men with azoospermia. *J Urol* **150**:1155-7.

Najmabadi H, Huang V, Yen P, Subbarao MN, Bhasin D, Banaag L, Naseeruddin S, et al (1996) Substantial prevalence of microdeletions of the Y-chromosome in infertile men with idiopathic azoospermia and oligozoospermia detected using a sequence-tagged site-based mapping strategy. *Journal of Clinical Endocrinology and Metabolism* **81**:1347-1352.

Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA Probe Detecting Multiple Haplotypes of the Human Y Chromosome. *American Journal of Human Genetics* **38**:407-418.

Niederberger C, Agulnik AI, Cho Y, Lamb D, Bishop CE (1997) In situ hybridization shows that *Dazla* expression in mouse testis is restricted to premeiotic stages IV-VI of spermatogenesis. *Mammalian Genome* **8**:277-278.

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti A, Dupanloup I, Langaney A, Excoffier L (1997a) Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Americal Journal of Human Genetics* **61**:1015-35.

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997b) Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong
correspondence with linguistics. *American Journal of Human Genetics* **61**:1015-35.

Prosser J, Inglis JD, Condie A, Ma K, Kerr S, Thakrar R, Taylor K, et al (1996) Degeneracy in human multicopy RBM (YRRM), a candidate spermatogenesis gene. *Mammalian Genome* **7**:835-42.

Qureshi SJ, Ross AR, Ma K, Cooke HJ, Intyre MA, Chandley AC, Hargreave TB (1996) Polymerase chain reaction screening for Y chromosome microdeletions: A first step towards the diagnosis of genetically-determined spermatogenic failure in men. *Molecular Human Reproduction* **2**:775-779.

Reijo R, Alagappan RK, Patrizio P, Page DC (1996a) Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome. *Lancet* **347**:1290-1293.

Reijo R, J. S, Dinulos MB, Jaffe T, Brown LG, Disteche CM, Page DC (1996b) 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.

Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, et al (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genetics* **10**:383-393.

Rice KA (1987) Estimating nucleotide sequence divergence from polymorphism in nucleotide sequences and restriction maps. University of Georgia, Athens, Georgia

Rice WR (1994) Degeneration of a nonrecombining chromosome. *Science* **263**:230-232.

Rice WR (1996) Evolution of the Y sex chromosome in animals. *BioScience* **46**:331-343.

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

Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, et al (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.

Silber SJ, Alagappan R, Brown LG, Page DC (1998) Y chromosome deletions in azoospermic and severely oligozoospermic men undergoing intracytoplasmic sperm injection after testicular sperm extraction. *Human Reproduction* **13**:3332-3337.

Stuppia L, Calabrese G, Franchi PG, Mingarelli R, Gatta V, Palka G, Dallapiccola B (1996a) Widening of a Y-chromosome interval-6 deletion transmitted from a father to his infertile son accounts for an oligozoospermia critical region distal to the RBM1 and DAZ genes [letter]. *American Journal of Human Genetics* **59**:1393-1395.

Stuppia L, Mastroprimiano G, Calabrese G, Peila R, Tenaglia R, Palka G (1996b) Microdeletions in interval 6 of the Y chromosome detected by STS-PCR in 6 of 33 patients with idiopathic oligo- or azoospermia. *Cytogenetics Cell Genetics* **72**:155-158.

Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, et al (1999) An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. *Nature Genetics* :in press.

Tiepolo L, Zuffardi O (1976) Localization of factors controlling spermatogenesis in the nonfluorescent portion of the Y chromosome long arm. *Human Genetics* **34**:119-124.

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research* **7**:996-1005.

Vergnaud G, Page DC, Simmler M, Brown L, Rouyer F, Noel B, Botstein D, et al (1986) A deletion map of the human Y chromosome based on DNA hybridization. *American Journal of Human Genetics* **38**:109-124.

Vogel T, Speed RM, Teague P, Cooke HJ (1999) Mice with Y chromosome deletion and reduced Rbm genes on a heterozygous Dazl1 null background mimic a human azoospermic factor phenotype. *Human Reproduction* **14**:3023-9.

Vogt P, 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. *Human Genetics* **89**:491-496.

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, et al (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics* **5**:933-943.

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

Wong J, Blanco P, Affara NA (1999) An exon map of the AZFc male infertility region of the human Y chromosome. *Mammalian Genome* **10**:57-61.

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, et al (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* **60**:1174-83.

CHAPTER 2

The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned

The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned

Richa Saxena[1], Laura G. Brown[1], Trevor Hawkins[2], Raaji K. Alagappan[1], Helen Skaletsky[1], Mary Pat Reeve[2], Renee Reijo[1], Steve Rozen[1,2], Mary Beth Dinulos[3], Christine M. Disteche[3] & David C. Page[1,2]

It is widely believed that most or all Y-chromosomal genes were once shared with the X chromosome. The *DAZ* gene is a candidate for the human Y-chromosomal *Azoospermia Factor* (*AZF*). We report multiple copies of *DAZ* (>99% identical in DNA sequence) clustered in the *AZF* region and a functional *DAZ homologue* (*DAZH*) on human chromosome 3. The entire gene family appears to be expressed in germ cells. Sequence analysis indicates that the Y-chromosomal *DAZ* cluster arose during primate evolution by (i) transposing the autosomal gene to the Y, (ii) amplifying and pruning exons within the transposed gene and (iii) amplifying the modified gene. These results challenge prevailing views of sex chromosome evolution, suggesting that acquisition of autosomal fertility genes is an important process in Y chromosome evolution.

XY sex chromosomes are found in a multitude of species throughout the animal kingdom. It is thought that XY chromosomes arose independently in many evolutionary lineages, in each case deriving from an ordinary autosomal pair. According to prevailing theories[1–4], once recombination between nascent X and Y chromosomes becomes restricted, the gene content of the Y chromosome declines steadily and inexorably. Translocation may occasionally add new autosomal material to both X and Y, in which case the process of Y degeneration begins anew. Degeneration of the Y is well documented in *Drosophila*[5,6] and has been shown to be an ongoing process even among mammals, which are generally considered to exhibit extreme differentiation of the X and Y chromosomes (K. Jegalian and D.C.P., in preparation). The few genes that persist on highly differentiated Y chromosomes are thought to be relics of this common ancestry with the X chromosome. According to this view, Y-chromosomal genes were once (or still are) shared with the X chromosome[1–4,7]. To the extent that the Y accumulates new DNA sequences independently of the X chromosome, these DNA sequences are thought to be primarily transposable elements whose chief functional consequence is to accelerate the degeneration of Y-borne genes[2,4,6].

Theories traceable to R.A. Fisher provide counterpoint to these purely degenerative theories of Y evolution. In 1931, Fisher hypothesized that, in early stages of differentiation from the X chromosome, incipient Y chromosomes would tend to accumulate alleles (at genes close to but distinct from sex determining gene[s]) that enhance male fitness but diminish female fitness[8]. Such 'sexually antagonistic' or 'male benefit' alleles have emerged on incipient Y chromosomes produced by experimental design in *Drosophila*[9]. Could it be that the Y chromosome, even after extreme differentiation from the X, would tend to acquire genes that promote male fitness? This speculation is consistent with Burgoyne, who has argued that the Y chromosome should accumulate genes that enhance spermatogenesis[10]. But in no case has the Y chromosome been shown to have acquired anew such a fertility factor. Indeed, in no animal has a differentiated Y chromosome been shown to have procured an autosomal gene during evolution, independent of the X chromosome. As described below, an unexpected opportunity to reconstruct just such an evolutionary event arose while studying the human Y chromosome's *Azoospermia Factor* (*AZF*).

In 1976, Tiepolo and Zuffardi reported *de novo* deletions of the distal half of Yq in four men with azoospermia (no sperm detected in semen), and on this basis they postulated the existence of one or more Yq genes critical for spermatogenesis[11]. In recent years, this *Azoospermia Factor* (*AZF*) hypothesis has been amply validated. Exploiting the availability of comprehensive, DNA-probe-based physical maps of the Y chromosome[12–14], investigators have reported many interstitial Yq deletions in infertile men[15–18]. In particular, overlapping *de novo* deletions within intervals 6D–6E of the Y chromosome[19] have been shown to cause at least 13% of cases of nonobstructive azoospermia — and some cases of severe oligospermia (low sperm count) as well[19,20]. Men with deletions of this region are infertile but otherwise healthy, suggesting that *AZF* is a 'pure male sterile' locus with no somatic function.

[1]*Howard Hughes Medical Institute and* [2]*Center for Genome Research, Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA* [3]*Department of Pathology, University of Washington, Seattle, Washington 98195, USA*

Correspondence should be addressed to D.C.P.

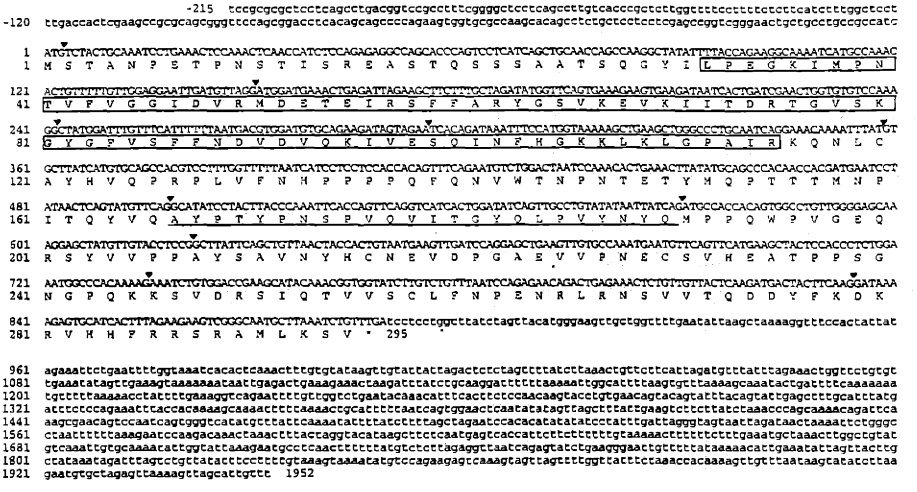


Fig. 1 Human *DAZH* cDNA sequence (clone pDP1648) and predicted amino acid sequence of encoded protein. RNP/RRM domain of protein is boxed. The single 24-amino-acid 'DAZ repeat' is underlined. Arrowheads above nucleotide sequence depict probable locations of ten introns (inferred by homology to *DAZ*; see Fig. 7). Numbering of nucleotides and amino acids begins with first in-frame AUG codon. GenBank accession number U65918.

The only transcription unit identified in this commonly deleted region is *DAZ* (*Deleted in Azoospermia*), a strong *AZF* candidate that encodes a putative RNA-binding protein[19]. Expression of *DAZ* is restricted to testes[19], where the gene is transcribed in premeiotic germ cells, particularly in spermatogonia, the earliest cells of the spermatogenic lineage[21]. Thus, *DAZ* may function in the first stages of spermatogenesis, or even earlier, in maintaining germ stem cell populations, and this could readily account for the spermatogenic defects caused by *AZF* deletions. There are no reports of *DAZ* point mutations in infertile human males. However, a close homologue of the *DAZ* gene has been described in *Drosophila*, and loss-of-function mutations in this fly homologue, *boule*, result in azoospermia while sparing the soma[22], much like human *AZF*. These genetic studies in *Drosophila* provide strong if indirect evidence that *DAZ* is *AZF* in humans.

In *Drosophila* the *DAZ* homologous gene *boule* is autosomal[22], as is the mouse *DAZ homologue* (*Dazh*, also known as *Dazla*)[23, 24]. If autosomal *DAZ* homologues are found in these other animals, perhaps they also occur in humans? Indeed, when hybridized to Southern blots of human genomic DNAs, *DAZ* cDNA probes detect not only male-specific, Y-chromosomal fragments but also a male–female common band that could represent a human autosomal homologue (see Fig. 5 of ref. 19). Is this putative autosomal homologue a functional gene or a pseudogene? What is its relationship to Y-chromosomal *DAZ*? These questions led us to explore what we now appreciate to be the *DAZ* gene family in humans and, ultimately, to reconstruct a chapter in the evolution of the human Y chromosome.

Expressed *DAZ* homologue on chromosome 3

We previously analysed *DAZ* cDNA clones, obtained from a human adult testis library, that unambiguously

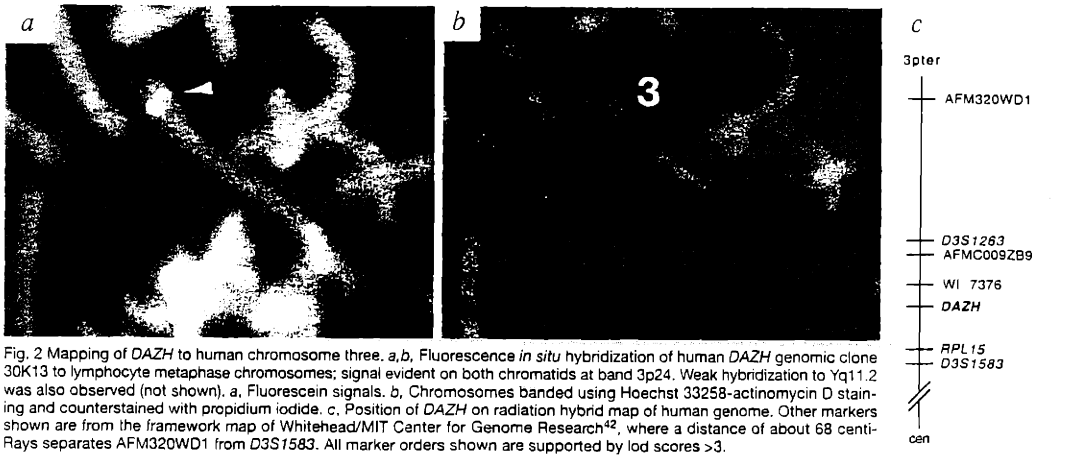


Fig. 2 Mapping of *DAZH* to human chromosome three. *a,b*, Fluorescence *in situ* hybridization of human *DAZH* genomic clone 30K13 to lymphocyte metaphase chromosomes; signal evident on both chromatids at band 3p24. Weak hybridization to Yq11.2 was also observed (not shown). *a*, Fluorescein signals. *b*, Chromosomes banded using Hoechst 33258-actinomycin D staining and counterstained with propidium iodide. *c*, Position of *DAZH* on radiation hybrid map of human genome. Other markers shown are from the framework map of Whitehead/MIT Center for Genome Research[42], where a distance of about 68 centiRays separates AFM320WD1 from *D3S1583*. All marker orders shown are supported by lod scores >3.

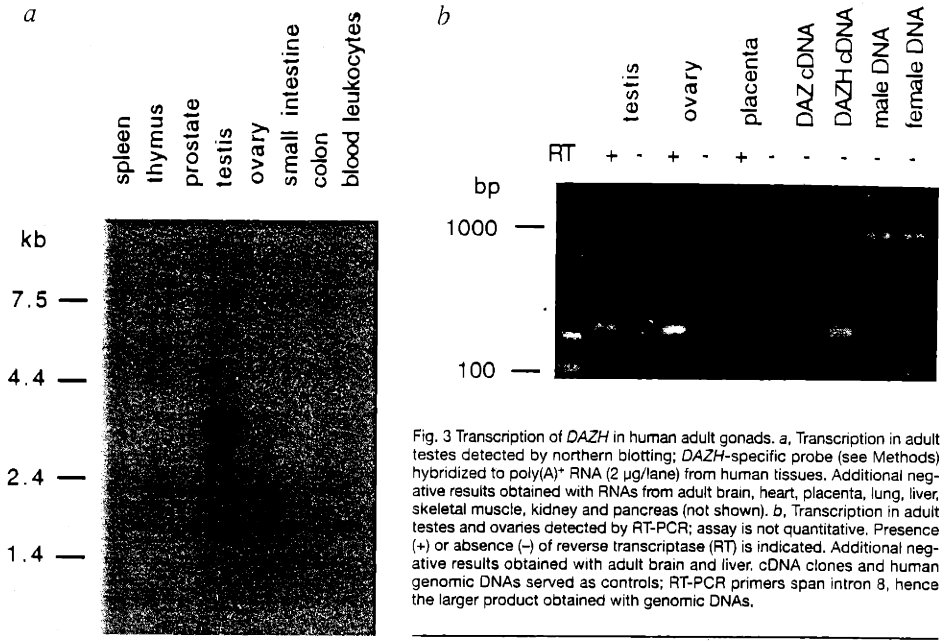


Fig. 3 Transcription of *DAZH* in human adult gonads. *a*, Transcription in adult testes detected by northern blotting; *DAZH*-specific probe (see Methods) hybridized to poly(A)$^+$ RNA (2 μg/lane) from human tissues. Additional negative results obtained with RNAs from adult brain, heart, placenta, lung, liver, skeletal muscle, kidney and pancreas (not shown). *b*, Transcription in adult testes and ovaries detected by RT-PCR; assay is not quantitative. Presence (+) or absence (–) of reverse transcriptase (RT) is indicated. Additional negative results obtained with adult brain and liver. cDNA clones and human genomic DNAs served as controls; RT-PCR primers span intron 8, hence the larger product obtained with genomic DNAs.

mapped to the *AZF* region of the Y chromosome[19]. Partial sequence analysis of other cDNA clones from the same library, identified by hybridization with *DAZ* probes, suggested that they were derived from a single transcription unit that was homologous but not identical to *DAZ*. We will refer to this homologous gene as *DAZH* (*DAZ homologue*). Complete sequence analysis of two *DAZH* cDNA clones revealed that they were collinear and shared a single long open reading frame (Fig. 1). This transcript appears to encode a protein of 295 amino acids, with a molecular weight of 33,170. As discussed below, the predicted DAZ and DAZH proteins are similar but nonidentical.

We then determined whether *DAZH*, like *DAZ*, mapped to the human Y chromosome. Using PCR assays specific to *DAZH*, we obtained products of identical size using human male or female genomic DNAs as templates, suggesting that the gene is autosomal or X-chromosomal. We mapped *DAZH* using two methods. By *in situ* hybridization of genomic BAC clones to human metaphase spreads, *DAZH* was localized to the distal short arm of chromosome 3 (band 3p24; Fig. 2*a*,*b*). This localization was independently confirmed and refined by PCR analysis of whole-genome radiation hybrid panels (Fig. 2*c*).

Human *DAZH* appeared to be expressed in adult testis, as indicated by our recovery of clones from a cDNA library prepared from this tissue. To confirm this result and to determine whether *DAZH* is transcribed elsewhere, we hybridized a *DAZH*-specific probe to northern blots of RNAs from 16 different human tissues. We also carried out RT-PCR analysis on five different human tissues using *DAZH*-specific primers. These studies revealed that *DAZH* is abundantly expressed in the adult testis, where a 3.5-kb transcript is readily detected by northern blotting (Fig. 3*a*), and is expressed at a lower level in the adult ovary, where a *DAZH*-specific RT-PCR product is observed (Fig. 3*b*). We detected no evidence of transcription in the other tissues examined.

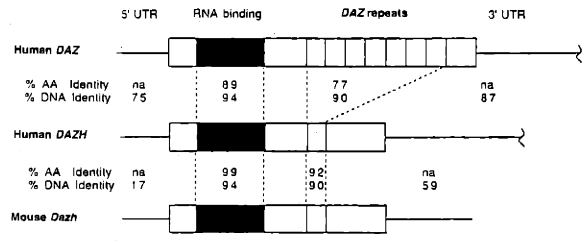


Fig. 4 Comparison of human *DAZ*, human *DAZH*, and mouse *Dazh* transcripts and encoded proteins[19,23,24]. This gene family encodes proteins with a single RNA-binding domain of the RRM/RNP type[43,44]. The human and mouse DAZH proteins have one copy of a 24-amino-acid unit that is tandemly repeated in DAZ. Percentage nucleotide and amino acid identities (na, not applicable) are shown for the following regions: 5' UTR, RNA binding domain, DAZ repeats, and 3' UTR.

The founding member of the *DAZ* gene family

Comparative analyses of predicted protein and underlying cDNA sequences for human *DAZH*, human *DAZ*, and mouse *Dazh* provided unexpected insights into the evolution of this gene family. The three proteins have quite similar structures, with overall sequence similarity being greatest between the products of the human *DAZH* and mouse *Dazh* genes (Fig. 4). Indeed, within the 82-residue RNA-binding domain, the products of human *DAZH* and mouse *Dazh*, both autosomal, differ by only one amino acid substitution, while both differ from human Y-encoded DAZ at nine residues. While the human Y-encoded protein includes seven tandemly arrayed 'DAZ repeats,' each 24 amino acids in length, the mouse and human DAZH proteins contain only one such unit.

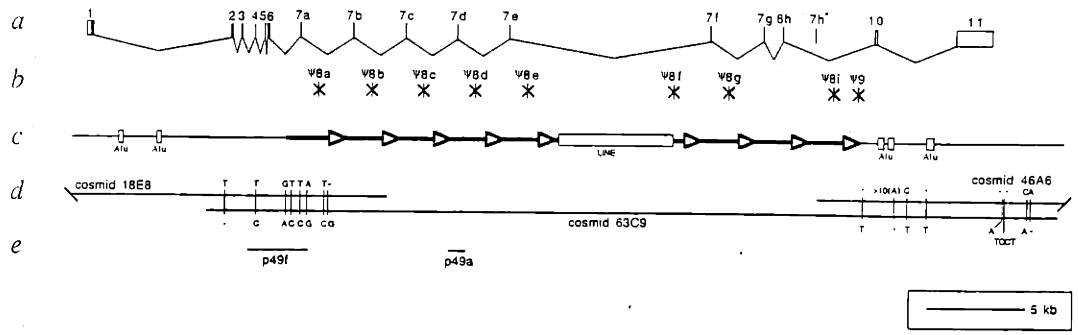


Fig. 5 Schematic representation of genomic DNA sequence from *DAZ* gene cluster on human Y chromosome. *a*, *DAZ* transcription unit. Exons numbered according to scheme outlined in Fig. 7; coding regions in black; UTRs in white. 7h, potentially an exon, has not been identified in sequenced cDNA clones (see text). *b*, Pseudoexons within *DAZ* transcription unit. *c*, Sequence backbone showing nine tandem repeats of a 2.4-kb unit, interrupted at one point by a 6.1-kb LINE element; Alu repeats indicated. *d*, Three cosmids from which sequence was derived. Nucleotide differences between 18E8 and overlapping portion of 63C9 or between 63C9 and overlapping portion of 46A6 are listed; deletions indicated by '–'. Sequence for both cosmids can be presently viewed at http://www-genome.wi.mit.edu; sequencing of 5' portion of cosmid 18E8 in progress. All of cosmid 46A6 (43,795 nucleotides) was sequenced, but only 12 kb is represented in the figure. *e*, Locations of *DYS1* plasmids p49f and p49a (refs 25, 30). Note: Vogt and colleagues have suggested that a second gene or gene family, designated *SPGY*, is found in the vicinity of *DAZ* in this *AZF* region. No sequence analysis of *SPGY* has been reported. However, Vogt and colleagues have reported two *SPGY* oligonucleotide sequences that yield a human genomic PCR product of 460 bp[18]. We find perfect matches to both oligonucleotides within *DAZ* exon 11 (nucleotides 8373–8398 and 8804–8829 in cosmid 46A6), where they span a region of 457 bp.

At first glance, these protein comparisons seemed to suggest that, during evolution, the ancestors of DAZ and DAZH diverged from a single common protein before the separation of the murine and human lineages. (In this case, the *DAZ* gene must have been lost or diverged beyond the point of cross-hybridization during murine evolution.) However, an examination of the cDNA sequences themselves clearly indicated a very different evolutionary course. Especially revealing were analyses of the genes' untranslated regions (UTRs), which are presumably subject to less intense selective pressures than are coding sequences and should evolve more rapidly. In their UTRs, the human *DAZH* and human *DAZ* transcripts exhibit a remarkably high degree of nucleotide sequence identity (75% and 87%, respectively, in 5' and 3' UTRs). A far lower degree of UTR sequence identity is observed between human *DAZH* and mouse *Dazh* (17% and 59%, respectively, for 5' and 3' UTRs). These UTR comparisons strongly suggested that the human *DAZ* and *DAZH* genes evolved from a single gene after, not before, the separation of murine and human lineages. This founding member of the human *DAZ* gene family must have encoded a protein much like human or mouse DAZH, given that the mouse (and fly) proteins show greater similarity to human DAZH than to human DAZ. (Homology between human chromosome 3, where *DAZH* maps, and mouse chromosome 17, where *Dazh*/*Dazla* maps[23,24], has not been reported previously.) On the basis of this analysis, we tentatively concluded that an ancestral, autosomal *DAZH* gene—still extant in humans, mice, and even flies—gave rise to Y-chromosomal *DAZ* during human evolution, after the separation of the human and murine lineages. The Y-encoded DAZ protein must have evolved relatively rapidly as compared with its highly conserved, autosomally encoded ancestor, DAZH. Subsequent analyses provided extensive corroboration of this model.

The *DAZ* gene cluster on the Y chromosome

To better understand the structure and evolution of the human *DAZ* gene family, we determined the nucleotide sequence of about 100 kb of the *AZF* region of the human Y chromosome (Fig. 5). The cosmids for sequence analysis were chosen, based on restriction fingerprinting and hybridization with *DAZ* oligonucleotides, to overlap modestly and to collectively span an entire *DAZ* transcription unit. The cosmids were derived from flow-sorted Y chromosomes originating from a single normal male. As we will describe, this sequence analysis confirmed our model of an autosome-to-Y transposition, revealed that the *DAZ* transcription unit had been shaped by an unprecedented process of exon amplification and pruning, and demonstrated that the *AZF* region contains multiple copies of *DAZ*. We detected no genes other than *DAZ* in the sequenced region.

Among the most evident features of the sequenced region is an array of nine tandem repeats of a 2.4-kb unit, comprising half of cosmid 63C9 (in the center in Fig. 5). These tandem repeats are interrupted at one point by a 6-kb LINE element, but they otherwise exhibit 77 to 96% sequence identity. As judged by numerous PCR assays on genomic DNAs from normal and *AZF*-deleted human males (not shown), these repeats appear to be specific to the *AZF* region of the Y chromosome.

The *DAZ* transcription unit appears to contain at least 16 exons and to span about 42 kb, including all nine tandem repeats. Located upstream of the 2.4-kb repeats are exon 1, which ends immediately 3' of the initiator codon, exons 2 through 5, which encode the RNA-binding domain, and exon 6. Each of the next seven exons (denoted 7a through 7g; see Fig. 7 for explanation of numbering system) is 72 bp in length, encodes a single 'DAZ repeat' of 24 amino acids, and falls within a 2.4-kb genomic repeat. Thus, seven of the first eight 2.4-kb tandem repeats appear to correspond, one to one, to the seven tandem 'DAZ repeats' previously noted in the encoded protein[19]. (The sixth tandem repeat is interrupted by the LINE element and lacks a 72-bp exon, apparently deleted at the site of the LINE's insertion.) Curiously, the subsequent exon (denoted exon 8) falls

within the eighth of the nine 2.4-kb tandem repeats, but its nucleotide and encoded amino acid sequences are unrelated to those of exons 7a–7g. The last two exons of *DAZ* are located 3' of the tandem repeat array. We have yet to identify a 3' poly(A)$^+$ tail in any *DAZ* cDNA clone. However, in the genomic DNA, a putative polyadenylation signal (AATAAA) is found 1.85 kb 3' of the 5' boundary of exon 11, and RT-PCR studies confirm that mature *DAZ* transcripts end shortly 3' of this polyadenylation signal.

Finally we compared in detail the three sequenced cosmids, all derived from a single individual's Y chromosome. We detected slight sequence differences among the three cosmids in regions of overlap, strongly suggesting that the cosmids represent distinct though highly similar copies of *DAZ*. Cosmids 18E8 and 63C9 appear to overlap by 8 kb (including exons 2 through 7b), but actually differ at eight nucleotides in this region (Fig. 5*d*). Similarly, cosmids 63C9 and 46A6 appear to overlap by 12 kb (including exons 10 and 11), but actually differ at eight sites (Fig. 5*d*). None of the nucleotide substitutions predicts an amino acid substitution or alters a splice site. As the three cosmids derive from a single individual, and thus a single Y chromosome, we cannot attribute these sequence differences to allelic variation but must instead conclude that they represent distinct copies of *DAZ* with approximately 99.9% sequence identity.

DYS1 is *DAZ*

We had previously reported[19] that the 72-bp repeat unit in the *DAZ* cDNA shows remarkable sequence similarity to human *DYS1*, an extraordinarily polymorphic family of Yq-specific sequences first described in 1984 and widely exploited since that time in population genetic studies[25–29]. A database search for DNA sequences related to the *DAZ* genomic locus revealed more extensive similarity to *DYS1*. We found near identity between the entirety of a sequenced segment (750 bp; plasmid p49a; ref. 30) of human *DYS1* and the fourth of the nine 2.4-kb repeats in *DAZ*.

These findings prompted us to examine more fully the relationship of *DYS1* to *DAZ* — and eventually to equate the two. First, we discovered the *Eco*RI restriction map of a *DYS1* cosmid (cosmid 49; Fig. 1 of ref. 26) to be strikingly similar to that of *DAZ* cosmid 63C9. Second, we found that PCR assays flanking *DAZ* exons 4, 5, 6, and 7a yielded products of the expected size when amplified from a *DYS1* clone (plasmid p49f; data not shown). As a final test of the equation, we probed Southern blots of *Taq*I-digested genomic DNAs from three *AZF*-deleted men (and their relatives) with plasmid p49f, the *DYS1* probe most widely employed in population genetic studies (Fig. 6). In normal male relatives, we observed the expected array of Y-specific *Taq*I fragments, both polymorphic and monomorphic. However, in the three *AZF*-deleted men, all Y-specific bands were absent, demonstrating that all *DYS1* sequences are, like the *DAZ* gene cluster, located in the *AZF* region. The only *DYS1*-homologous fragments remaining in the *AZF*-deleted men are two autosomal fragments (bands K and L in Fig. 6) that correspond to *DAZH* (as confirmed by *Taq*I digestion of *DAZH* BAC clones; data not shown). We conclude that the *DAZ* gene cluster and the highly polymorphic *DYS1* sequences are one and the same. In 1986,

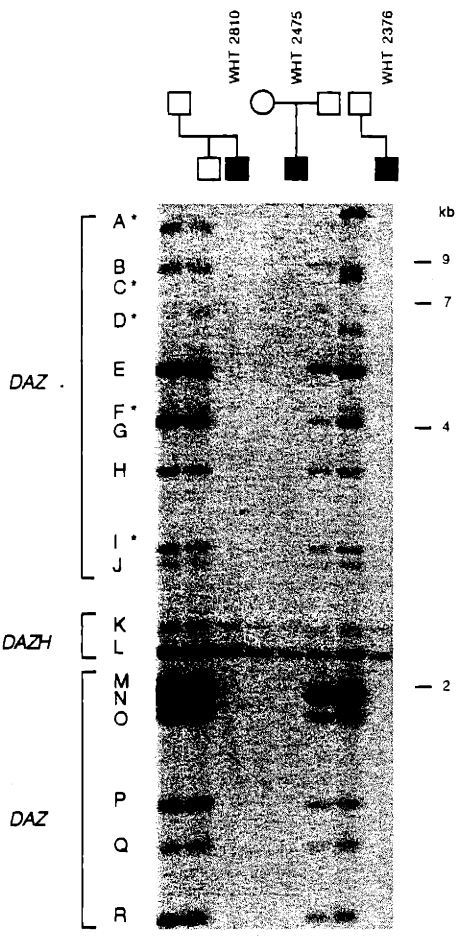


Fig. 6 *DYS1* probe p49f hybridized to Southern blot of *Taq*I-digested genomic DNAs (5 µg/lane) from three azoospermic men with *de novo* deletions of *AZF* region (and their immediate relatives; ref. 19 and R.A., Robert Oates, D.C.P, unpublished results). By convention[26], *Taq*I fragments hybridizing with p49f are labelled A through R. Fragments known to be polymorphic are indicated by an asterisk. Note differences between fathers of WHT2475 and WHT2376 in sizes of some Y-specific fragments. All Y-specific fragments are absent in each of the three *AZF*-deleted men (but present in their fathers) and correspond to the *DAZ* gene cluster. Fragments K and L, present in all individuals tested, correspond to *DAZH*. Scale in kb shown at right.

Seboun and colleagues[31] observed that *DYS1* was homologous to a testis-expressed gene on human chromosome 3 (evidently *DAZH*).

A transcription unit littered with vestigial exons

The *DAZH* coding region (Fig. 1) exhibited about 90% nucleotide sequence identity to the sequenced portion of the *AZF* region (Fig. 5), allowing us to deduce the likely locations of all *DAZH* introns (Figs 1,7) and to further explore the evolutionary relationship of the Y-chromosomal *DAZ* and autosomal *DAZH* transcription units. This analysis dramatically substantiated what we already suspected: while the *DAZH* gene appears to have a conventional structure, the *DAZ* transcription unit is a contorted derivative littered with degenerate

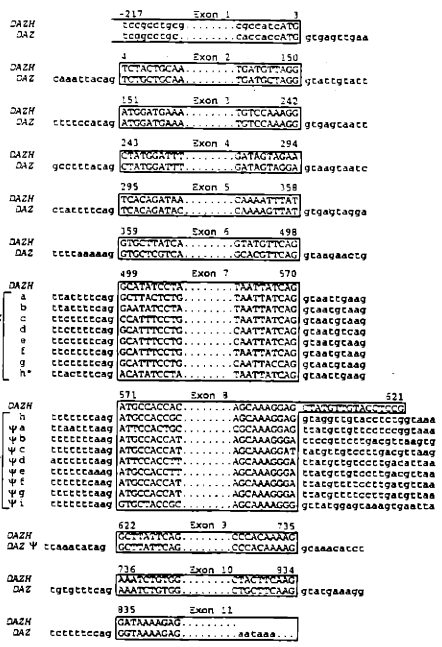


Fig. 7 Exons and pseudoexons of the human *DAZ* and *DAZH* genes. The figure is arranged in 11 tiers corresponding to the 11 exons of *DAZH* (boundaries inferred by homology to *DAZ*). In each tier the *DAZH* exon is shown in the top line, and below are shown all homologous regions, both exons and pseudoexons, in *DAZ* cosmids 18E8 (exon 1 through pseudoexon 8b) and 63C9 (exon 2 through exon 11; see Fig. 5). *DAZH* translated sequences (and homologous portions of *DAZ*) are capitalized. *DAZ* exon 8h is 16 nucleotides shorter at its 3' end than *DAZH* exon 8, apparently because a single nucleotide substitution created a new splice donor site in *DAZ*. *As described in text, 7h may be a true exon but has not been observed in cDNA clones.

exons. Indeed, scattered among the exons of a single *DAZ* transcription unit (largely encompassed by cosmid 63C9) are nine sequence segments that bear unmistakable similarity to *DAZH* exons yet consist of nothing more than vestigial remains of those exons. We will refer to these degenerate, vestigial exons as 'pseudoexons,' by analogy to 'pseudogenes.' Eight of the nine pseudoexons are relics of *DAZH* exon 8 and are found in the 2.4-kb tandem repeats that comprise the central half of the *DAZ* transcription unit. The remaining pseudoexon (a descendant of *DAZH* exon 9) is found between the last of the 2.4-kb repeats and exon 10. All nine *DAZ* pseudoexons share two properties that distinguish them from true *DAZ* exons. First, their 5' or 3' splice sites have degenerated (Fig. 7). Second, we have not found these pseudoexons in any of the *DAZ* cDNA clones we have sequenced, suggesting that they are excised (as components of introns) during processing of *DAZ* transcripts. The exon 7 derivative within the last of the 2.4-kb repeats (h in Fig. 7) may represent a tenth pseudoexon, as we have not detected it in any of the *DAZ* cDNA clones sequenced (ref. 19; data not shown), though its splice sites appear to be intact.

Discussion

Transposition, amplification and pruning. An examination of all available sequence information for the human Y-chromosomal *DAZ* and autosomal *DAZH* genes, cDNAs, and their encoded proteins suggests the following sequence of evolutionary events:

1. Transposition. A complete copy of the *DAZH* transcription unit was transposed from an autosome (what is now human chromosome 3) to the Y chromosome during primate evolution. This transposition occurred sometime prior to the splitting of the orangutan and human lineages, as indicated by the presence of male-specific, *DAZ*-homologous sequences in both species (see Fig. 5 in ref. 19).

2. Expansion and pruning of the transcription unit. Within the newly transposed gene, a 2.4-kb genomic segment encompassing exons 7 and 8 was tandemly amplified, eventuating in a long array such as that observed in cosmid 63C9. But in most of the amplified units, one or both of the exons degenerated or was deleted. For example, early in the course of the amplification process, a repeat unit arose in which exon 8 had been incapacitated by splice site mutations or other degenerative changes, and subsequent amplification of this unit gave rise to the present string of 2.4-kb repeats harbouring a functional derivative of *DAZH* exon 7 and a vestige of *DAZH* exon 8. Only in the penultimate repeat were both exons 7 and 8 preserved. The transposed descendant of *DAZH* exon 9 degenerated without amplification. With this one exception, the pruned *DAZ* transcription unit retained one or more functional descendants of each *DAZH* exon.

3. Gene amplification. The emerging *DAZ* transcription unit, having undergone internal duplications and substantial pruning, was amplified so that small numbers of transcription units exist in close proximity in the *AZF* region of the human Y chromosome. Our present data provide direct evidence for the existence of at least two or three copies of *DAZ* exhibiting 99.9% sequence identity (two if nonoverlapping cosmids 18E8 and 46A6 derive from the same copy of *DAZ*). This is a minimum estimate of gene copy number; the true number of *DAZ* copies may be greater. Indeed, when either *DAZ* or *DAZH* probes are hybridized to human genomic Southern blots, the resulting male-specific *DAZ* bands are far more intense than the male–female-common *DAZH* bands, even though the *DAZH* gene is present in two copies per cell, unlike the Y chromosome (Fig. 6; see also Fig. 5 of ref. 19).

Given the well documented polymorphism of the synonymous *DYS1* sequence family, we should anticipate that the sequence of some *DAZ* gene copies may be more diverged, at least in some individuals. Indeed, we observed 11 nucleotide differences between *DAZ* cDNA[19] (GenBank U21663) and genomic sequences, eight of these differences being in exons 7d and 7e. These differences could reflect sequence divergence among *DAZ* gene copies on a single Y chromosome, or they could reflect true polymorphisms that distinguish the individuals from whom cDNA and genomic libraries were prepared.

Preservation of function. The *DAZ* gene cluster on the human Y chromosome arose from an autosomal ancestor, *DAZH*, via a series of structural transformations whose complexity could not have been anticipated. Nonetheless, it appears that the newly emergent Y gene cluster retained key functional characteristics of its auto-

somal ancestor. First, the sequence of the encoded protein was largely preserved. The products of *DAZ* and *DAZH* appear to be RNA-binding proteins whose sequences are, apart from the 24-residue tandem repeats in DAZ, quite similar throughout much of their lengths. Such preservation of the bulk of the mature transcript's reading frame is a remarkable outcome given that the *DAZ* transcription unit encompasses 26 exons and pseudoexons, as compared with 11 exons in *DAZH*.

Second, it appears that both the ancestral and the more recently derived members of the *DAZ* gene family are expressed exclusively in germ cells. Like its mouse homologue (*Dazh*/*Dazla*; refs 23,24), human *DAZH* is abundantly transcribed in adult testes and at a lower level in adult ovaries (Fig. 3), while human *DAZ*, absent in females, is transcribed exclusively in testes[19]. As demonstrated by the absence of transcripts in germ-cell deficient mice (*White-spotted* and *Steel* mutants), *Dazh* expression in testes is restricted to germ cells[23], and we have recently extended these mutant studies to ovaries, with identical results (J. Seligman, R.R., D.C.P., unpublished results). In the adult human testis, the *DAZ* gene family is transcribed in spermatogonia and perhaps also in early spermatocytes, as revealed by *in situ* hybridization studies[21]. Thus, in both humans and mice, germ cells appear to be the only site of expression of the *DAZ* gene family.

It seems likely that the products of the ancestral gene, autosomal *DAZH*, and its derivative, Y-chromosomal *DAZ*, interact with similar or identical RNA targets in the same cell types. The similar azoospermic phenotypes associated with human *DAZ* deletions[19] and with loss-of-function mutations in the *Drosophila* homologue[22] suggest that the germ cell functions of the DAZ protein family may have been conserved throughout much of metazoan evolution. In humans, partial redundancy of Y-chromosomal *DAZ* and autosomal *DAZH* function could contribute to the variable nature of the spermatogenic defects caused by *AZF* deletions[19, 20]. Conversely, mutations in *DAZH* could be responsible for spermatogenic defects in some men with intact Y chromosomes.

Evolution of the Y chromosome. The case of human *DAZ* challenges the prevailing view[1–4,7] that most if not all Y-chromosomal genes were once shared with the X chromosome. We strongly affirm that much of the gene content of the Y chromosome reflects the Y's common ancestry and ongoing meiotic and functional relationship with the X. A substantial fraction of human Y chromosomal genes and DNA sequences have X homologues[13,14,32]. However, our results suggest that the Y chromosome's evolution and gene content may also be influenced by a process that is independent of the X chromosome. We speculate that the direct acquisition of autosomal genes that enhance male fertility is an important component of Y chromosome evolution. Selective pressures would favour this process, particularly if the genes transposed to the Y were of little or no benefit to females, and most especially if they diminished female fitness[1,2,4,8,9,33–35].

DAZ represents the first unambiguous example of autosome-to-Y transposition of a germ-cell factor, but diverse observations suggest that there may be other cases. Several other genes or gene families on the human, mouse or *Drosophila* Y chromosomes are expressed specifically in testes, where they likely function in spermatogenesis, and exhibit no evidence of X homology[32,36]. Could some of these genes have autosomal ancestors? Though not definitive, these observations suggest the possibility that autosome-to-Y transposition of male fertility factors may be a recurrent theme in Y chromosome evolution.

Regardless of chromosomal origin, genes transposed to the nonrecombining portion of the Y chromosome would inevitably face and likely succumb to powerful degenerative forces during subsequent evolution[1,2,4,7]. Perhaps the rate of acquisition of male fertility genes approximates the rate of subsequent degeneration, resulting in an evolutionary steady state. In contrast to the extreme evolutionary stability of the X chromosome, at least in mammals[3,37,38], individual male fertility genes might not be long-lived, in an evolutionary sense, on the Y chromosome.

Methods

***DAZH*-specific PCR assay.** A single pair of primers, one located in *DAZH* exon 8 (5'–GGAGCTATGTTGTACCTCC–3') and the other in *DAZH* exon 9 (5'–GTGGGCCATTTCCAGAGGG–3'), was used in PCR screening of a BAC, in typing of radiation hybrids, and in RT-PCR assays. These primers yield a 128-bp product from *DAZH* cDNA clones and a 0.8-kb product from human genomic DNA (Fig. 3*b*). This assay does not co-amplify *DAZ* genomic or cDNA sequences (Fig 3*b*); in *DAZ*, the homologue of *DAZH* exon 9 is a pseudoexon (Fig. 7). PCR was performed in 20 µl volumes of 1.5 mM $MgCl_2$, 5 mM NH_4Cl, 10 mM Tris-HCl (pH8.3), 50 mM KCl, 100 µM dNTPs, with 1 U *Taq* DNA polymerase and 1 µM of each primer. Thermocycling conditions: initial denaturation of 3 min at 94 °C; 35 cycles of 1 min at 94 °C, 1.5 min at 56 °C,1 min at 72 °C; and, finally, 5 min at 72 °C. RT-PCR (cDNA cycle kit, Invitrogen) was performed on 100 ng of total RNA from each of five human tissues (Clontech).

Chromosomal fluorescence *in situ* hybridization. *DAZH* clone 30K13 was isolated from the human genomic BAC library of Shizuya *et al.*[39](Research Genetics) by PCR screening. This BAC library was labelled with biotin-11 dATP by nick translation (Gibco BRL). Metaphase chromosomes were prepared from human male lymphocytes using 75 mM KCl as hypotonic buffer and methanol/acetic acid (3:1 v/v) as fixative. Hybridization was carried out as described[40] and signals were detected using a commercial system (Vector). The slides were blocked with goat serum, incubated with fluorescein avidin DCS, and rinsed in 4× SSC, 0.03% Triton. Slides were then incubated with biotinylated anti-avidin D and rinsed again. A second incubation with fluorescein avidin DCS was followed by a final rinse. Chromosomes were banded using Hoechst 33258-actinomycin D staining and counterstained with propidium iodide. Chromosomes and hybridization signals were visualized by fluorescence microscopy using a dual band pass filter (Omega).

Radiation hybrid mapping. DNAs from the 93 hybrid cell lines of the GeneBridge 4 panel[41] (Research Genetics) were tested for *DAZH* by PCR. Analysis of the results unambiguously positioned *DAZH* with respect to the radiation hybrid framework map constructed at the Whitehead/MIT Center for Genome Research[42].

Northern and Southern blotting. A *DAZH*-specific hybridization probe was derived from *DAZH* cDNA clone pDP1648 by PCR using the primers described above. This probe, labelled by incorporation of [^{32}P]-dCTP during PCR, was hybridized overnight to northern blots of human tissue RNAs (Fig. 3*a*; Clontech) at 65 °C in 1 M sodium phosphate (pH 7.5), 7% SDS.

Blots were washed three times for 20 min each at 37 °C in 0.1× SSC, 0.1% SDS. For Southern blotting (Fig. 6), the purified insert of *DYS1* plasmid p49f (ref. 25) was [^{32}P]-labelled by random-primed synthesis and hybridized overnight using the conditions just described, except that blots were washed at 42 °C in 2× SSC, 0.1% SDS.

Genomic DNA sequencing. *AZF*-region cosmids were selected[19] from a Y-enriched library (LL0YNC03) constructed at the Human Genome Center, Lawrence Livermore National Laboratory, Livermore, CA. A complete description of the methods employed in sequencing cosmids 63C9, 46A6 and 18E8 will be presented elsewhere (T.L.H. and colleagues, in preparation). Briefly, M13 and pUC libraries were prepared from each cosmid, and standard dye-primer based shotgun sequencing methods were used to obtain six-fold coverage, on average, of the cosmid insert. The sequence was completed using primer-directed chemistries and directed reverse reads. Further information on the sequencing project can be found at http://www-genome.wi.mit.edu.

GenBank accession numbers. *DAZ*: U21663; DAZH: U65918; cosmid 63C9 and cosmid 46A6: pending.

Acknowledgements

We thank the members of the DNA Sequencing Group at the Whitehead/MIT Center for Genome Research for their efforts in sequencing the three cosmids described; J. Seligman and G. Mutter for human RNAs; S. Silber and R. Oates for patient samples; Lawrence Livermore National Laboratory for the flow-sorted cosmid library; and W. Rice, K. Jegalian, B. Lahn, and B. Charlesworth for helpful discussions and comments on the manuscript. Supported by National Institutes of Health, Howard Hughes Medical Institute, and March of Dimes Birth Defects Foundation; genomic sequencing supported by grants from NIH/NCHGR and DOE to T.L.H.; R.R. was recipient of a Damon-Runyon/Walter Winchell fellowship.

Received 12 August; accepted 20 September 1996.

1. Bull, J.J. *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, Menlo Park, California, 1983).
2. Charlesworth, B. The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**,149–162 (1996).
3. Ohno, S. *Sex Chromosomes and Sex-Linked genes* (Springer Verlag, Berlin, 1967).
4. Rice, W.R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
5. Rice, W.R. Degeneration of a nonrecombining chromosome. *Science* **263**, 230–232 (1994).
6. Steinemann, M. & Steinemann, S. Degenerating Y chromosome of *Drosophila miranda*: a trap for retroposons. *Proc. Natl. Acad. Sci. USA* **89**, 7591–7595 (1992).
7. Graves, J.A.M. The origin and function of the mammalian Y chromosome and Y-borne genes — an evolving understanding. *BioEssays* **17**, 311–321 (1995).
8. Fisher, R.A. The evolution of dominance. *Biol. Rev.* **6**, 345–368 (1931).
9. Rice, W.R. Sexually antagonistic genes: experimental evidence. *Science* **256**, 1436–1439 (1992).
10. Burgoyne, P.S. Fruit(less) flies provide a clue. *Nature* **381**, 740–741 (1996).
11. Tiepolo, L. & Zuffardi, O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the Y chromosome long arm. *Hum. Genet.* **34**, 119–124 (1976).
12. Vergnaud, G. *et al.* A deletion map of the human Y chromosome based on DNA hybridization. *Am. J. Hum. Genet.* **38**, 109–124 (1986).
13. Vollrath, D. *et al.* The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* **258**, 52–59 (1992).
14. Foote, S., Vollrath, D., Hilton, A. & Page, D.C. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* **258**, 60–66 (1992).
15. Johnson, M.D., Tho, S.P.T., Behzadian, A. & McDonough, P.G. Molecular scanning of Yq11 (interval 6) in men with Sertoli-cell-only syndrome. *Am. J. Obstet. Gynecol.* **161**, 1732–1737 (1989).
16. Skare, J. *et al.* Interstitial deletion involving most of Yq. *Am. J. Med. Genet.* **36**, 394–397 (1990).
17. Ma, K. *et al.* Towards the molecular localisation of the *AZF* locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. *Hum. Molec. Genet.* **1**, 29–33 (1992).
18. Vogt, P.H. *et al.* Human Y chromosome *Azoospermia Factors* (*AZF*) mapped to different subregions in Yq11. *Hum. Molec. Genet.* **5**, 933–943 (1996).
19. Reijo, R. *et al.*. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genet.* **10**, 383–393 (1995).
20. Reijo, R., Alagappan, R.K., Patrizio, P. & Page, D.C. Severe oligospermia resulting from deletions of *Azoospermia Factor* gene on Y chromosome. *The Lancet* **347**, 1290–1293 (1996).
21. Menke, D., Mutter, G. & Page, D.C. Expression of *DAZ*, an *Azoospermia Factor* candidate, in human spermatogonia. *Am. J. Hum. Genet.* (in the press).
22. Eberhart, C.G., Maines, J.Z. & Wasserman, S.A. Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. *Nature* **381**, 783–785 (1996).
23. Reijo, R. *et al.* 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 (1996).
24. Cooke, H.J. *et al.* A murine homologue of the human *DAZ* gene is autosomal and expressed only in male and female gonads. *Hum. Mol. Genet.* **5**, 513–516 (1996).
25. Bishop, C., Guellaen, G., Geldwerth, D., Fellous, M. & Weissenbach, J. Extensive sequence homologies between Y and other chromosomes. *J. Mol. Biol.* **173**, 403–417 (1984).
26. Ngo, K.Y., Vergnaud, G., Johnsson, C., Lucotte, G. & Weissenbach, J. A DNA probe detecting multiple haplotypes of the human Y chromosome. *Am. J. Hum. Genet.* **38**, 407–418 (1986).
27. Lucotte, G., Guerin, P., Halle, L., Loirat, F. & Hazout, S. Y chromosome DNA polymorphisms in two African populations. *Am. J. Hum. Genet.* **45**, 16–20 (1989).
28. Santachiara Benerecetti, A.S. *et al.* The common, Near Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. *Ann. Hum. Genet.* **57**, 55–64 (1993).
29. Spurdle, A. & Jenkins, T. Y chromosome probe 49a detects complex PvuII haplotypes and many new TaqI haplotypes in southern African populations. *Am. J. Hum. Genet.* **50**, 107–125 (1992).
30. Lucotte, G., David, F. & Mariotti, M. Nucleotide sequence of p49a, a genomic Y-specific probe with potential utilization in sex determination. *Mol. Cell. Probes* **5**, 359–363 (1991).
31. Seboun, E. *et al.* A molecular approach to the study of the human Y chromosome and anomalies of sex determination in man. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 237–248 (1986).
32. Affara, N. *et al.* Report of the second international workshop on Y chromosome mapping 1995. *Cytogenet. Cell Genet.* **73**, 33–76 (1996).
33. Winge, O. The location of eighteen genes in *Lebistes reticulatus*. *J. Genet.* **18**, 1–43 (1927).
34. Charlesworth, D. & Charlesworth, B. Sex-differences in fitness and selection for centric fusions between sex chromosomes and autosomes. *Genet. Res.* **35**, 205–214 (1980).
35. Page, D.C. Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads. *Development* **101 Suppl.**, 151–155 (1987).
36. Hackstein, J.H. & Hochstenbach, R. The elusive fertility genes of *Drosophila*: the ultimate haven for selfish genetic elements. *Trends Genet.* **11**, 195–200 (1995).
37. Rugarli, E. *et al.* Different chromosomal localization of the *Clcn4* gene in *Mus spretus* and C57BL/6J mice. *Nature Genet.* **10**, 466–471 (1995).
38. Palmer, S., Perry, J. & Ashworth, A. A contravention of Ohno's law in mice. *Nature Genet.* **10**, 472–476 (1995).
39. Shizuya, H.B. *et al.* 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 (1992).
40. Chance, P.F. *et al.* DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* **72**, 143–151 (1993).
41. Gyapay, G. *et al.* A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**, 339–346 (1996).
42. Hudson, T.J. *et al.* An STS-based map of the human genome. *Science* **270**, 1945–1954 (1995).
43. Burd, C.G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA binding proteins. *Science* **265**, 615–621 (1994).
44. Kenan, D.J., Query, C.C. & Keene, J.D. RNA recognition: towards identifying determinants of specificity. *Trends Biochem.* **16**, 214–220 (1991).

CHAPTER 3

Four *DAZ* genes found in two clusters within the *AZFc* region of the human Y chromosome

(adapted from a manuscript in preparation for submission, with *Genomics* format)

Four *DAZ* genes in two clusters found in *AZFc* region of human Y chromosome

Richa Saxena, Jan W. A. de Vries,* Sjoerd Repping,† Raaji K. Alagappan, Helen Skaletsky, Laura G. Brown, Peter Ma,‡ Ellson Chen,‡ Jan M. N. Hoovers,† and David C. Page

Howard Hughes Medical Institute, Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

*Center for Reproductive Medicine, Department of Gynecology and Obstetrics, and †Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands

‡Celera Genomics, Foster City, CA 94404, USA

ABSTRACT

The *DAZ* genes are candidate fertility factors that lie within the human Y chromosome's *AZFc* region, whose deletion is a common cause of spermatogenic failure. The number of *DAZ* genes has been difficult to determine, in part because the nucleotide sequences of the *DAZ* genes are nearly identical. Here, fluorescence in situ hybridization and characterization of BAC clones revealed four full-length *DAZ* genes on the human Y chromosome. They exist in two clusters, each comprising an inverted pair of *DAZ* genes (3'←5'::5'→3'). Analysis of genomic sequences and testicular transcripts suggested that three or four *DAZ* genes are translated. Each gene contains at least seven tandem copies of a previously described, 2.4-kb repeat unit that encodes 24 amino acids. In addition, two *DAZ* genes contain tandem copies of a 10.8-kb repeat unit that encodes the RNA binding domain, which appears to be multimerized in some DAZ proteins. Combining our present results with previous studies, we can reconstruct several steps in the evolution of the *DAZ* genes on the Y chromosome. In the ancestral Y-chromosomal *DAZ* gene, amplification of both intragenic repeats began before the human and cynomolgus (Old World) monkey lineages diverged. During subsequent evolution, an inverted duplication of this modified gene occurred. Finally, the resulting two-gene cluster was duplicated, generating the two-cluster/four-gene arrangement found on modern human Y chromosomes.

Key Words: Y chromosome, male infertility, DAZ, AZFc, testis

INTRODUCTION

Approximately two percent of men are infertile because they produce few or no sperm (Hull *et al.*, 1985; Silber, 1989). The most common known molecular cause of such spermatogenic failure is deletion of the *AZFc* region on the long arm of the human Y chromosome (Ma *et al.*, 1992; Reijo *et al.*, 1995; Vogt *et al.*, 1996). Although one or more spermatogenesis genes must lie within the *AZFc* region, the identity of the critical factor(s) is still uncertain because no point mutations or internal deletions in candidate genes have been identified. Candidate genes within this region include *DAZ, BPY2* , *RBMY*, and *CDY1* (Reijo *et al.*, 1995; Lahn and Page, 1997; Yen, 1998).

The *DAZ* (*Deleted in Azoospermia*) genes, which encode putative RNA-binding proteins, are strong *AZFc* candidates. The *DAZ* genes are located exclusively within the *AZFc* region and are transcribed only in testicular germ cells (Reijo *et al.*, 1995; Saxena *et al.*, 1996; Menke *et al.*, 1997). In model organisms, genetic studies have demonstrated that *DAZ* homologs play essential roles in germ cell development (Eberhart *et al.*, 1996; Ruggiu *et al.*, 1997; Houston and King, 2000). In mice, disruption of the *Dazl* gene leads to germ cell loss before birth, rendering both males and females infertile (Ruggiu *et al.*, 1997). In *Drosophila*, males mutant for the *DAZ* homolog *boule* are infertile with germ cell arrest at the G2/M transition into meiosis I (Eberhart *et al.*, 1996).

The precise number of *DAZ* genes in the *AZFc* region has been difficult to determine. Initially, only one *DAZ* gene was thought to exist within the *AZFc* region (Reijo *et al.*, 1995). We later reported that *DAZ* cosmids derived from a single individual differed slightly in DNA sequence, providing evidence for at least two distinct *DAZ* genes (Saxena *et al.*, 1996). Glaser and colleagues found evidence of multiple *DAZ* genes on Yq by fluorescence in situ hybridization (Glaser *et al.*, 1997). By Southern blotting and long-range restriction mapping, Yen and colleagues obtained evidence of at least three *DAZ* genes on the Y chromosome (Yen *et al.*, 1997; Yen, 1998). Most recently, using fiber-FISH, Glaser, Yen and Schempp observed what they interpreted to be seven *DAZ* genes or pseudogenes on Yq (Glaser *et al.*, 1998).

How many of the *DAZ* genes on the Y chromosome are functional? Are some of them pseudogenes? Genes on Y chromosomes are often subject to degeneration during evolution (Ohno, 1967; Rice, 1994; Charlesworth, 1996). Repetitive gene families on the human Y

chromosome may include both functional and corrupted gene copies. For example, *RBMY* is present in at least ten copies throughout the Y chromosome, but only a few copies appear to be transcribed and functional (Elliott *et al.*, 1997). Similarly, the *TSPY* gene family on the human Y chromosome consists of a mixture of active and silent gene copies (Manz *et al.*, 1993). By analogy, one might anticipate that the Y-linked *DAZ* genes would include some transcriptionally active and some decayed family members. Yen and colleagues found a variety of *DAZ* transcripts in single individuals. These findings raise the possibility that multiple *DAZ* genes are expressed, but the transcripts could also have arisen by alternative splicing of a single gene (Yen *et al.*, 1997). If only one *DAZ* gene was shown to be expressed, then one might search that gene for point mutations in men with spermatogenic failure to identify a critical role in spermatogenesis. If multiple *DAZ* genes were expressed, it would be less likely that spermatogenic failure would be caused by *DAZ* point mutations.

We sought to build upon previous studies and to address several questions. Exactly how many *DAZ* genes with an intact genomic structure are present on human Y chromosomes? How many of these genes are transcribed and translated? How are the genes arranged, and how do they differ from each other? What can be inferred about the pathway by which the *DAZ* genes evolved on the human Y chromosome?

MATERIALS AND METHODS

Cosmids. We used three sequenced *DAZ* cosmids in these studies. Cosmid 18E8 (Saxena *et al.*, 1996) has an insert of 42,791 bp, corresponding to nucleotides 670 through 43,460 in the recently sequenced BAC RP11-290O13 (Genbank AC010089). As shown in Fig. 1, cosmid 18E8 encompasses the 5' portions of two neighboring *DAZ* genes. Cosmid 63C9 ((Saxena *et al.*, 1996); Genbank AC000021) contains exons 2-11 and thus almost an entire *DAZ* gene. Cosmid 46A6 ((Saxena *et al.*, 1996); Genbank AC000022) derives from the 3' portion of *DAZ*; it contains exons 8-11 as well as 35 kb downstream of the gene.

Fluorescence in situ hybridization (FISH). One or two-color FISH was performed according to standard procedures (Redeker *et al.* 1994). Probes were labeled with biotin or

digoxigenin, hybridized to target DNA, and detected by avidin or anti-digoxigenin antibodies conjugated to fluorochromes Cy3 (red) or fluorescein (green).

Extended chromatin fibers from spermatozoa were prepared as described previously (Haaf and Ward, 1995) with minor modifications. Sperm were isolated by density centrifugation on a 70% Percoll gradient, washed twice in phosphate-buffered saline (PBS), resuspended in a 3:1 mixture of methanol/acetic acid to 10^7 sperm/ml, allowed to fix for 1 h at –20°C, and dropped onto glass slides. After blow drying, slides were incubated in extraction solution (0.125% SDS, 0.2M NaOH) for 5 min at 30°C. The solution was removed, and new extraction solution was pipetted onto one end of the slide and smeared out using a coverslip. This procedure was repeated using fixative (3:1 methanol/acetic acid). The slides were dehydrated and kept at room temperature prior to hybridization.

Extended chromatin fibers from lymphocytes were prepared using SDS/EDTA extraction (Fidlerova *et al.*, 1994).

BACs. *DAZ* BACs were isolated from human male genomic libraries prepared at the California Institute of Technology (Shizuya *et al.*, 1992). We probed high-density library filters (Research Genetics) using radiolabeled PCR products corresponding to *DAZ* STSs. A total of 16 *DAZ* BACs were identified. Three BACs (prefixed with CTA) derive from DNA of one male donor. The remaining 13 BACs (prefixed with CTB) derive from a second, unrelated male donor. BAC DNA was isolated using alkaline lysis and column chromatography (Qiagen) using pre-heated elution buffer.

Pulsed-field gel electrophoresis. *DAZ* BACs were sized by pulsed-field gel electrophoresis in 1% agarose using a Bio-rad CHEF-DRII system. Electrophoresis was performed for 26 h at 15°C and 179 V with ramped switch times of 5 to 20 s. Estimated BAC sizes (including vector sequences) were as follows: CTA-50D17, 240 kb; CTA-132B16, 122 kb; CTA-148I14, 110 kb; CTB-235I11, 165kb; CTB-263M7, 130 kb; CTB-293A20, 170 kb; CTB-315F14,140 kb; CTB-327P21, 130 kb; CTB-352E14, 200 kb; CTB-374C1, 100 kb; CTB-387E18, 138 kb; CTB-415B11, 160 kb; CTB-482K23, 175 kb; CTB-492O16, 200kb; CTB-530K16, 150 kb; and CTB-546E5, 135 kb.

For Southern analysis of *DAZ* genes, restriction-digested BACs were subjected to electrophoresis for 11 h at 14°C and 200 V with ramped switch times of 1 to 6 s. This separated restriction fragments ranging in size from 5 to 75 kb.

Southern blotting. Following agarose gel electrophoresis, restriction-digested BAC and cosmid DNAs were transferred onto Genescreen Plus (NEN) membranes and hybridized with radiolabeled *DAZ* PCR products or plasmid insert (pDP 1649; 2.4 kb insert from *DAZ* genomic locus). Probes were labeled with ^{32}P-dCTP by random priming. Hybridization was carried out at 65° C in 0.5 M $NaPO_4$ (pH7.5), 7% SDS. Membranes were subsequently washed at 65°C in 0.1X SSC, 0.1% SDS three times for 20 min each.

Detection of Sequence Family Variants (SFVs) that distinguish between DAZ genes. PCR amplification was performed in 20 μl volumes of 1.5 mM $MgCl_2$, 5 mM NH_4Cl, 10 mM Tris (pH8.3), 50 mM KCl, 100 μM dNTPs, with 1 U Taq DNA polymerase and 1μM of each primer. PCR primers and conditions are deposited in Genbank: sY581, Genbank G63906; sY586, G63907; sY587, G63908; sY579, G63909; sY776, G63910. To detect SFVs at sY581, sY586 and sY587, PCR products were digested with restriction enzymes as listed in Table 1.

cDNA cloning and sequencing. DAZ cDNA clones were identified by screening a library (HL1161X, Clontech) prepared from testes of four men; the screening methods were described previously (Reijo *et al.*, 1995). Lambda phage cDNA clones were converted into pDR plasmids (pDP1575, pDP1576, pDP1678, pDP1679), or their inserts were PCR amplified and subcloned into pBluescript plasmids (pDP1680 and pDP1681, with overlapping inserts together representing a single isolate from the cDNA library).

Because of lengthy tandem repeats, *DAZ* cDNA clones were not amenable to nucleotide sequencing by conventional methods. Instead, we sequenced from transposon inserts into cDNA subclones (Devine *et al.*, 1997). Briefly, for cDNA clones pDP1575, pDP1678, pDP1679 and pDP1680, a library of recombinant plasmids carrying transposon insertions were prepared using a Primer Island Transposition Kit (PE Applied Biosystems) *in vitro*. The transposition reaction was terminated by adding freshly prepared stop buffer (0.25 M EDTA, 1% SDS, 5mg/ml Proteinase K) and incubating at 65°C for 30 min. Excess reagents were removed by precipitating

the products with isopropanol solution (25 μl water, 25 μl 7.5 M ammonium acetate, 75 μl isopropanol) and washing with 70% ethanol.

The resulting plasmid DNAs were electroporated (Gene Pulser; BIO-RAD) into DH10B *E. coli* cells (Life technology) at a setting of 25 uFD, 200 ohm, 2.5 V. Subsequent sample preparation and DNA sequencing were carried out as described (Chen *et al.*, 1996), employing primers PIP (3'-CAGGACATTGGATGCTGAGAATTCG-5') and PIM (3'-CAGGAGCCGTCTATCCTGCTTGC-5') with BigDye (PE Applied Biosystems) terminator chemistry. Sequence data were assembled using computer software *Phred/Phrap* and edited using *Consed* (http://www.phrap.org).

RESULTS

Two Clusters, Each Containing an Inverted Pair of DAZ Genes

Previously, we reported the partial sequence of 5' *DAZ* cosmid 18E8, which together with cosmids 63C9 and 46A6 provided the composite sequence of a *DAZ* transcription unit (Saxena *et al.*, 1996). Further sequencing and mapping of cosmid 18E8 revealed a nearly perfect inverted duplication comprising most of the cosmid's insert (Fig. 1). One arm of the inverted sequence contains *DAZ* exons 1 through 7d. The other arm, which extends to the cosmid's cloning site, contains a second copy of exon 1 (and part of intron 1) in reverse orientation. A non-duplicated segment of 2.1 kb (including a THE element) lies between the inverted repeats.

This sequencing of cosmid 18E8 suggested that at least one inverted pair of *DAZ* genes might exist on the Y chromosome. We corroborated and extended this model through fluorescence *in situ* hybridization (FISH) analysis. We hybridized *DAZ* cosmid probes to human male chromatin in three different states of condensation: 1) in interphase fibroblast nuclei, 2) in extended chromatin fibers from spermatozoa and 3) in fully extended chromatin fibers from lymphocytes.

Representative results of hybridizing *DAZ* cosmids to interphase fibroblast nuclei are shown in Fig 2a,b. Cosmid 18E8 (5' *DAZ*) generated two signals in 75% of nuclei examined (Fig. 2a). In the remaining nuclei (25%), one signal was observed, likely from the superimposition of the two signals. By contrast, 3' *DAZ* cosmid 46A6 (Fig. 2b) produced four

signals in 41% of nuclei examined, with the remaining nuclei exhibiting three (28%), two (24%), or one signal (7%). Superimposition of signals may account for the nuclei exhibiting three or fewer signals. These findings suggested 1) that there are four *DAZ* genes on the Y chromosome and 2) that the 5' ends of the *DAZ* genes (two FISH signals) are in closer proximity than their 3' ends (up to four FISH signals), consistent with head-to-head *DAZ* gene duplication (3'←5'::5'→3').

To achieve higher resolution, we hybridized *DAZ* cosmids to extended chromatin fibers from spermatozoa. There, two-color FISH with *DAZ* cosmids 63C9 and 46A6 revealed two large signal clusters (Fig. 2c). Within each cluster, the 46A6 signal (3' *DAZ*; green) overlaps the outer ends of the 63C9 signal (central portion of *DAZ*; red), as expected if two head-to-head *DAZ* genes are present in each cluster. We repeated these studies on six other unrelated men, in each case observing the same pattern of two clusters, with evidence of 3'←5'::5'→3' orientation within each cluster.

To examine the orientation of *DAZ* genes within a cluster in detail, we performed two-color FISH on extended chromatin fibers from lymphocytes of two men. We consistently observed evidence that each cluster contained two *DAZ* genes in head-to-head orientation (Figure 2d, e). For example, Fig. 2d shows two separate 46A6 signals (3'*DAZ*; green) flanking one continuous 18E8 signal (5' *DAZ*; red). Similarly, Fig. 2e shows 46A6 signals (3' DAZ; red) overlapping the outer ends of two long 63C9 signals (central portion of *DAZ*; green).

Taken together, our FISH studies suggested that human Y chromosomes carry two *DAZ* clusters, each containing two *DAZ* genes in 3'←5'::5'→3' orientation.

Sequence Family Variants that Distinguish Between DAZ1, DAZ2, DAZ3, and DAZ4

To scrutinize the hypothesis of four *DAZ* genes in two clusters, and to compare the genes in detail, we used *DAZ* probes to screen human male BAC libraries providing an estimated 4-to-5-fold coverage of the Y chromosome. We identified and characterized 16 *DAZ* BAC clones. A physical map of the four *DAZ* genes based on studies of these BAC clones is presented in Fig. 3. We will now describe the analyses that enabled us to construct this physical map.

Our previous studies of *DAZ* cosmids revealed that the *DAZ* genes are >99.9% identical in DNA sequence (Saxena *et al.*, 1996). These previous studies had also identified a handful of

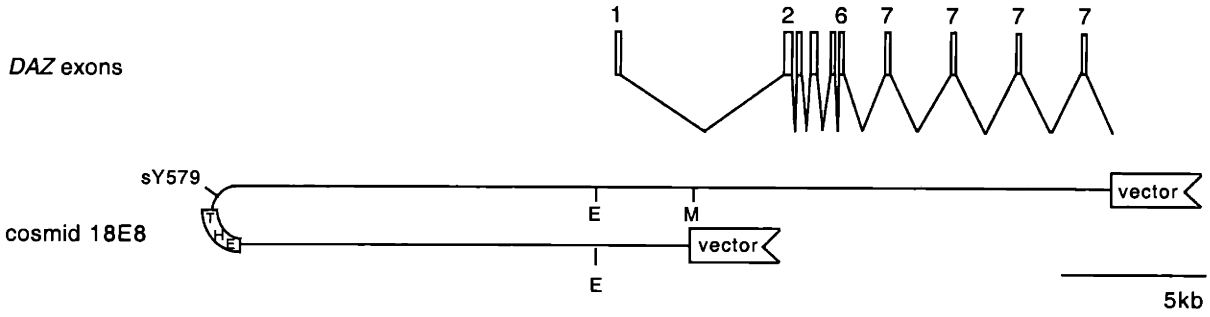


Figure 1. Schematic diagram of sequence from cosmid 18E8. *DAZ* exons are shown above. Positions of a 1.9kb THE element and STS marker sY579 in between two inverted *DAZ* genes are indicated. Cosmid 18E8 sequence corresponds to nucleotide 672 through 43,460 in sequence of unfinished BAC RP11-290O3 (Genbank AC010089). Restriction sites for restriction enzymes *Eag*I (E) and *Mlu*I (M) are indicated.

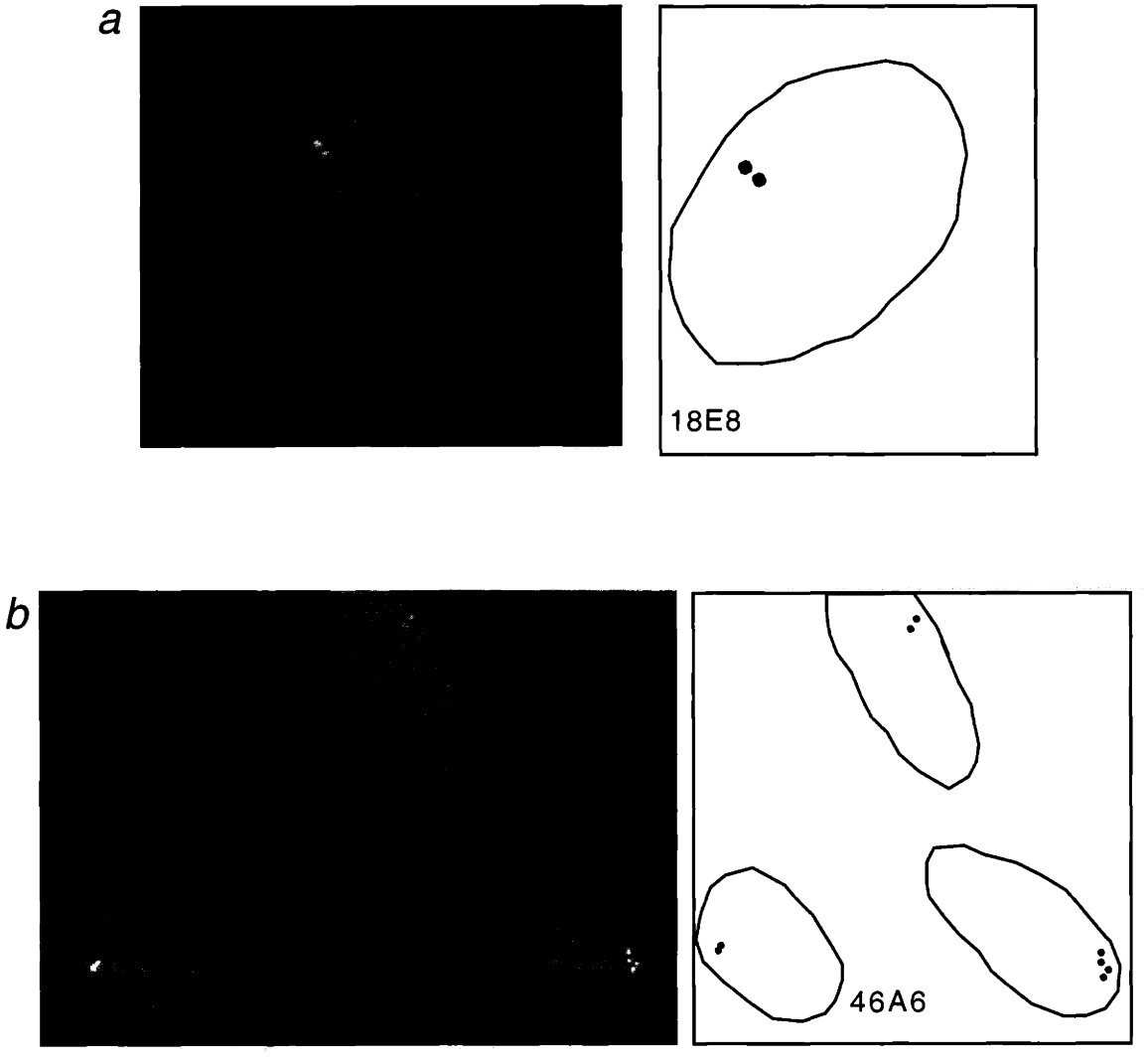


Figure 2*a,b* Fluorescence in situ hybridization using a)DAZ cosmid 18E8 (5' end of DAZ with exons 1-7d, 44kb insert) and b)cosmid 46A6 (3' end of DAZ with exons 8-11, 43.8 kb insert) on interphase nuclei from human fibroblasts.

Figure 2 *c* Two-color FISH with cosmid probes 63C9 (almost entire gene with exons 2-11,41.5kb insert; red) and 46A6 (green) on extended DNA fibers from spermatozoa.

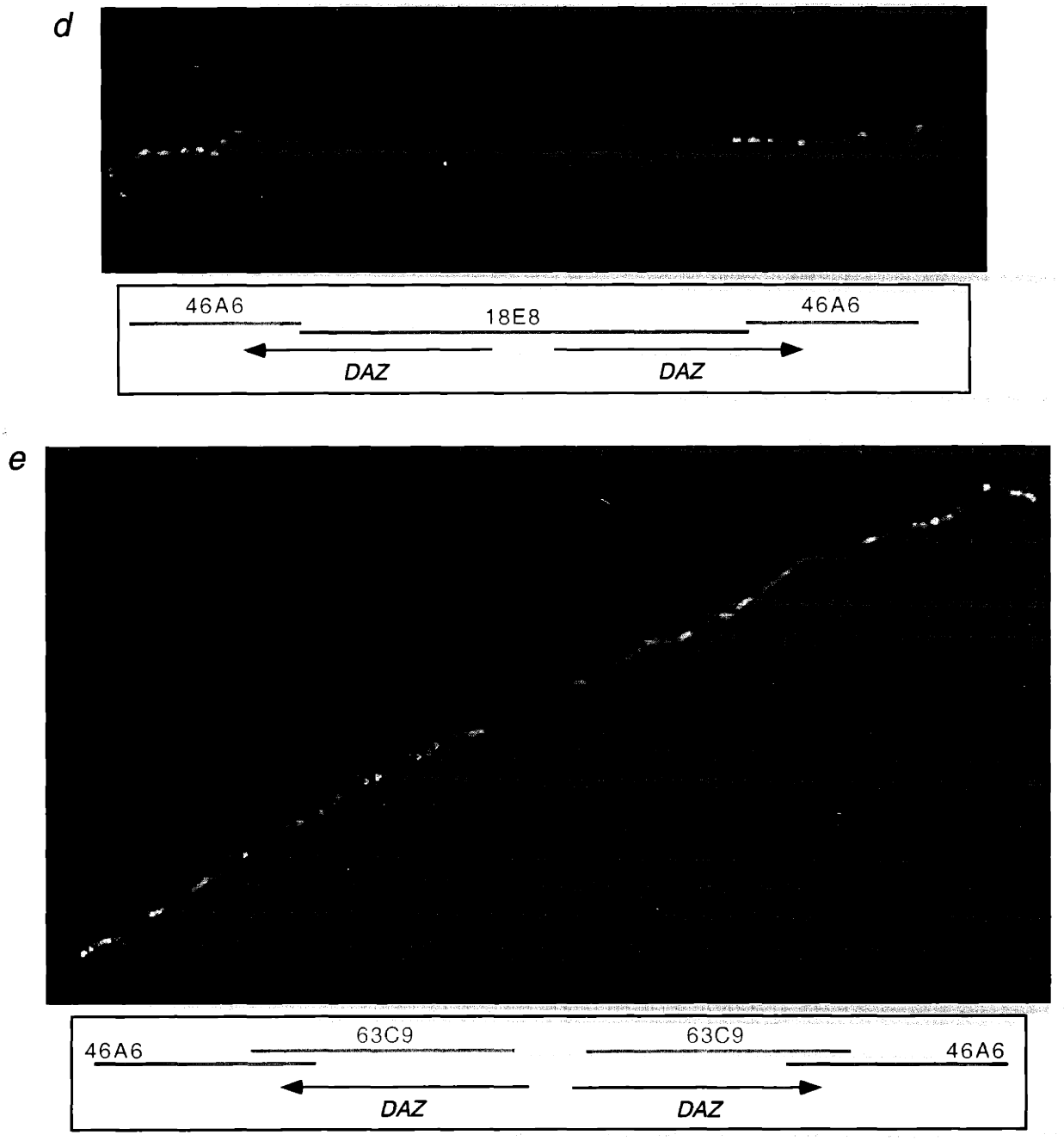


Figure 2 *d,e* Two-color FISH on extended DNA from lymphocytes d)probes 18E8 (red), 46A6 (green) e)probes 63C9 (green), 46A6 (red).

single nucleotide variants that might be useful, we reasoned, in distinguishing among the *DAZ* genes. We developed PCR/restriction-digest assays to type the BACs for these single nucleotide variants. Typing of the 16 BACs for three sequence variants (sY581/Sau3A, sY586/TaqI, and sY587/DraI) revealed four distinct *DAZ* gene signatures — *DAZ1*, *DAZ2*, *DAZ3* and *DAZ4* (see Table 1 and Fig. 4 for details). Nine of the 16 BACs exhibited a single signature — either *DAZ1*, *DAZ2*, *DAZ3* or *DAZ4* — consistent with each carrying a single *DAZ* gene.

The seven other BACs exhibited two signatures each — either *DAZ1* plus *DAZ2*, or *DAZ3* plus *DAZ4*. We hypothesized that each of these seven BACs contained portions or all of two *DAZ* genes in a head-to-head orientation like that found in cosmid 18E8. Several pieces of evidence corroborated this interpretation. Each of the two-signature BACs contained sY579, an STS located between the 5' ends of the inverted *DAZ* genes found in cosmid 18E8 (Fig. 1). Similarly, restriction digestion and pulsed-field gel electrophoresis of these seven BACs revealed that each contained an *Eag*I fragment of 20 kb (data not shown), as also seen in 5' cosmid 18E8 (Fig.1). The apparent pairing of *DAZ1* with *DAZ2* (in BAC CTB-235I11), and of *DAZ3* with *DAZ4* (in six independent BAC clones), suggested the precise composition of the two *DAZ* clusters visualized by FISH. *DAZ1* or *DAZ2* were never seen in the same BAC clone as *DAZ3* or *DAZ4*, consistent with the *DAZ1/2* and *DAZ3/4* clusters being too far apart for both clusters to be captured within a BAC insert.

Intact Coding Sequences in DAZ1, DAZ2, DAZ3, and DAZ4

The single nucleotide variants that we had used to distinguish among the *DAZ* genes were all located in introns. Having identified BACs corresponding to each of the four *DAZ* genes, we then compared the genes' coding regions at the nucleotide level. For each of the four genes, we sequenced exons 1 through 7a (the 5'-most copy of exon 7; Saxena et al., 1996) and exons 8 through 11, using BACs as sources of sequencing templates. As judged by this limited genomic sequence analysis, the coding regions of all four genes appeared to be intact, with no evidence of frameshift or nonsense mutations in *DAZ1*, *DAZ2*, *DAZ3*, or *DAZ4*. Indeed, we observed only one coding sequence difference among the *DAZ* genes: a silent C-to-T transition in exon 7a in *DAZ2*.

Tandem Amplification of 10.8-kb Unit within DAZ1 and DAZ3

We then compared the four *DAZ* genes at a structural level by additional restriction mapping of their respective BACs. Conventional and pulsed-field Southern blotting of BAC DNAs

enabled us to identify restriction fragments of particular interest. Hybridization probes employed in these studies included PCR products and synthetic oligonucleotides corresponding to specific exons, as well as plasmid subclones of portions of the genes. The resulting maps and inferred arrangements of exons are summarized in Fig. 3, where, in the interest of clarity, only selected restriction sites are shown.

This restriction mapping/Southern blot analysis of *DAZ* BACs yielded several insights. First, the four *DAZ* genes differ in size, as revealed most directly by pulsed-field gels following digestion with *PmeI*, which cuts near the 5' and 3' ends of all four genes. The approximate sizes of the genes are as follows: *DAZ1*, 65 kb; *DAZ2*, 70 kb; *DAZ3*, 55 kb and *DAZ4*, 50 kb.

Our analysis of *DAZ* BACs also revealed that, in the central portions of all four genes, there are tandem arrays of a previously identified 2.4-kb unit. Previous sequencing of *DAZ1* cosmid 63C9 (Saxena *et al.*, 1996) had identified this genomic repeat and revealed that it contains a 72-bp exon (exon 7) encoding a 24-amino acid segment that is tandemly amplified within predicted DAZ proteins (Reijo *et al.*, 1995; Yen *et al.*, 1997). As shown in Fig. 5a, hybridization of a 2.4-kb-repeat probe to restriction-digested BAC DNAs revealed a set of large fragments — similar to those seen in cosmid 63C9 — in each of the four genes. These and other Southern blot analyses of *DAZ* BACs indicated that the 2.4-kb unit is tandemly amplified in all four genes. As summarized in Fig. 3, all four *DAZ* genes appear to contain many copies of exon 7.

Finally, our analysis of *DAZ* BAC clones revealed a second tandemly amplified segment within *DAZ* genes: a 10.8-kb unit that is triplicated in *DAZ1* and duplicated in *DAZ3*, as summarized in Fig. 5c. Nucleotide sequence analysis of *DAZ* cosmids previously revealed only two *Mlu*I restriction sites within a composite *DAZ* transcription unit — one site in intron 1 and another site in one copy of exon 7 (Saxena *et al.*, 1996). We hybridized a genomic probe encompassing exons 2 and 3 to pulse-field Southern blots of *Mlu*I-digested BAC DNAs, expecting to observe one hybridizing fragment per gene. Indeed, as shown in Fig. 5b, we observed a single hybridizing fragment in BACs containing either *DAZ2* (BAC CTB-352E14) or

Table 1

Sequence Family Variants that Distinguish Between DAZ Genes Detected by PCR/Restriction-Digest Analysis

STS Marker	PCR Product Size (bp)	Restriction Endonuclease	Fragment Sizes (bp) after Restriction Digests			
			DAZ1	*DAZ2*	*DAZ3*	*DAZ4*
sY581 (intron 3)	252	*Sau*3A	189, 63	130,59, 63	189, 63	130,59, 63
sY586 (intron 6)	301	*Taq*I	184,117	301	184,117	184,117
sY587 (intron 10)	270	*Dra*I	122,73, 49,26	122,73 49,26	195, 49,26	195, 49,26

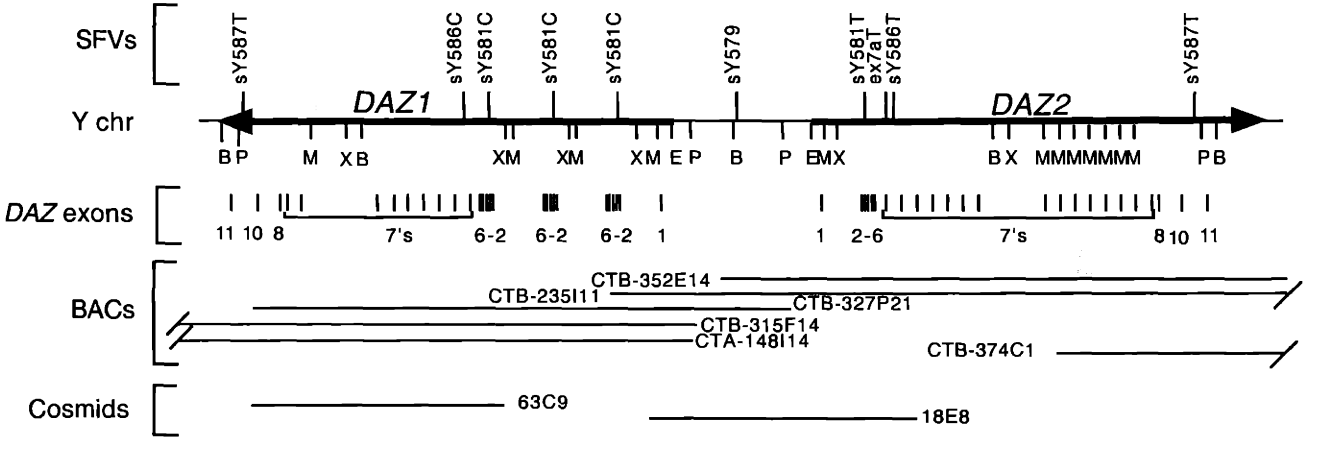


Figure 3a. Genomic organization of one cluster of *DAZ* genes(*DAZ1,DAZ2*) inferred from analysis of recombinant clones. STSs detecting single nucleotide variants between *DAZ* genes and STS sY579 are shown above the line depicting the Y chromosome. Each *DAZ* gene is labeled and indicated with a large arrow. A restriction map with key features of each *DAZ* gene is shown below(B *BamH*I, P *Pme*I, M *Mlu*I, X *Xho*I, E *Eag*I). *DAZ* exons are shown. BAC and cosmid clones used for this analysis are shown as horizontal bars below the map, only portions containing *DAZ* genes are shown. CTA-132B16 and BAC CTA-148I14 comes from DNA from a different individual than other BAC clones.

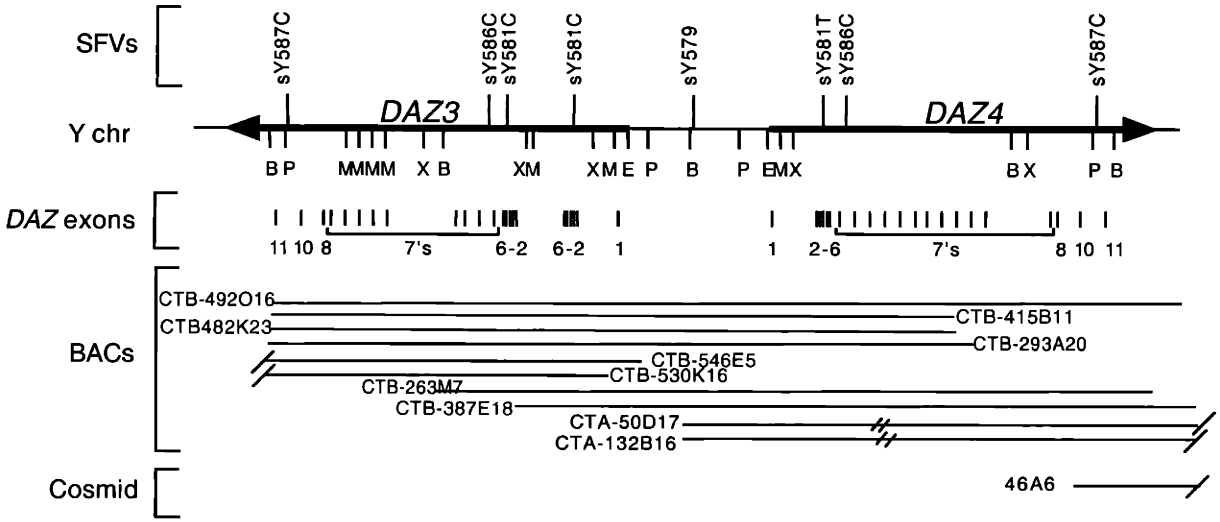


Figure 3b. Genomic organization of one cluster of *DAZ* genes (*DAZ3,DAZ4*) inferred from analysis of recombinant clones. STSs detecting single nucleotide variants(SFVs) between *DAZ* genes and STS sY579 are shown above the line depicting the Y chromosome. Each *DAZ* gene is labeled and indicated with a large arrow. A restriction map with key features of each *DAZ* gene is shown below(B *BamH*I, P *Pme*I, M *Mlu*I, X *Xho*I, E *Eag*I). *DAZ* exons are shown. BAC and cosmid clones studied are shown as horizontal bars below the map, only portions within them that contain *DAZ* genes are shown. BACs CTA-50D17, CTA-132B16 come from DNA from a different individual than other BAC clones. Slanted lines in the middle of clones CTA-50D17 and CTA-132B16 containing *DAZ4* are to indicate that these clones have two more 2.4kb repeats before the LINE than other *DAZ4* containing clones.

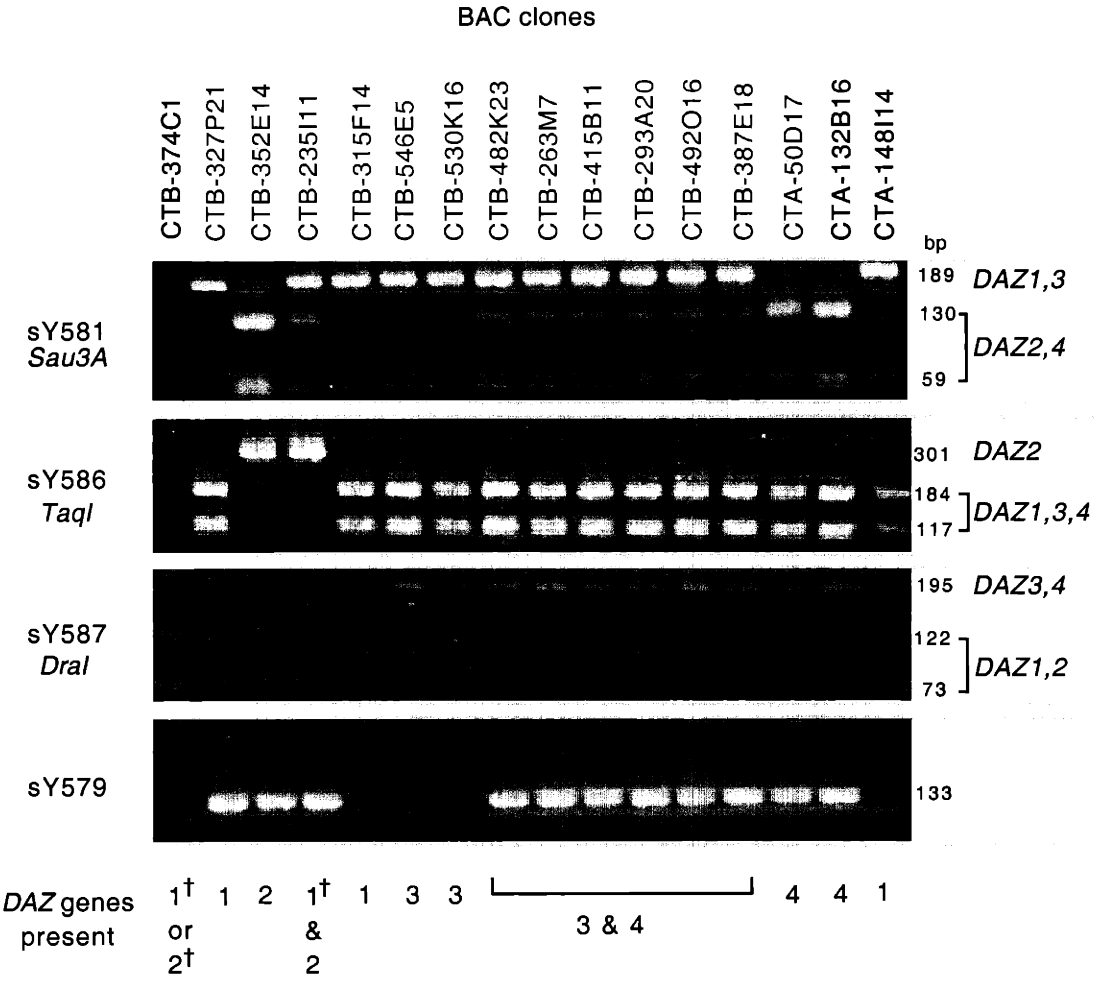


Figure 4. EtBr stained gel of PCR amplified DNA digested with restriction enzymes to detect sequence family variants between *DAZ* BAC clones. PCR-restriction digest assays sY581,sY586 and sY587 are described in Table 1, and their positions within *DAZ* genes are indicated on the map in Figure 3. Fragment sizes in bp are indicated to the right. Also listed are the DAZ genes each fragment may arise from. STS sY579 detects sequence in between inverted DAZ genes (see Figure1).

DAZ4 (BAC CTA-132B16). However, we observed three or two hybridizing *Mlu*I fragments in BACs containing *DAZ1* (BACs CTA-148I14 and CTB-327P21) or *DAZ3* (BAC CTB-546E5), respectively. These results suggested that *DAZ1* and *DAZ3* contained, respectively, three and two copies of exons 2-3. Additional Southern-blot studies of BAC DNAs revealed that exons 4-6 are also present three times in *DAZ1* and twice in *DAZ3* (data not shown). These findings were corroborated and extended by nucleotide sequence analysis of *DAZ1* BAC CTA-148I14 (GenBank AF164343; David Schwartz and colleagues, personal communication), which revealed a tandem triplication of a 10.8-kb genomic unit encompassing exons 2 through 6.

In 1992, Vollrath and colleagues identified sY152, an STS mapping to deletion interval 6D (Vollrath *et al.,* 1992). This STS was subsequently shown to be deleted in infertile men with *AZFc/DAZ* deletions (Reijo *et al.*, 1995). However, Kent-First and colleagues concluded recently 1) that sY152 is located proximal to both *AZFc* and the *DAZ* gene cluster, and 2) that sY152 defines a novel region (which they have referred to as "*AZFd* ") whose deletion causes spermatogenic failure (Kent-First *et al.,* 1999). We find, as depicted in Fig. 5c, that sY152 is located within the 10.8-kb repeats in the *DAZ1* and *DAZ3* genes. These findings confirm that sY152 is an *AZFc/DAZ* marker, contradicting the conclusions of Kent-First and colleagues.

Expression of Three or Four DAZ Genes in the Testis

As described above, our analysis of *DAZ* genomic sequences suggested that the coding sequences of all four *DAZ* genes were intact. However, genomic sequencing alone could not reveal whether each of the four genes was transcribed in vivo. To examine this question, we sequenced a variety of *DAZ* cDNA clones and sought to assign them to individual *DAZ* genes.

We first isolated 17 *DAZ* cDNA clones from a human testis cDNA library made from RNAs pooled from four individuals. We selected the five longest clones and sequenced them in their entirety. Sequencing of *DAZ* cDNA clones is difficult because of lengthy tandem repeats within the coding regions, and few if any *DAZ* cDNA clones had been fully and accurately sequenced in previous studies (see discussion in Yen *et al.,* 1997). To circumvent these difficulties, we inserted transposons into the cDNA clones, thereby introducing unique priming sites for sequencing. Three of the five sequenced cDNA clones appeared to be full-length, containing a complete, intact *DAZ* open reading frame. By comparing cDNA and genomic

sequences, we were able to assign the first of the full-length cDNA clones to *DAZ2*, the second to *DAZ4*, and the third to *DAZ3* or *DAZ1*.

This nucleotide sequence analysis allowed us to predict the primary structures of the DAZ proteins, which are depicted schematically, together with the autosomally encoded DAZL protein (Saxena *et al.*, 1996), in Fig. 6. The 24-amino-acid units that are tandemly repeated in DAZ proteins show some variability in sequence, as recognized previously (Reijo *et al.*, 1995; Yen *et al.*, 1997). To denote the distinct forms of the 24-amino-acid repeat (encoded by distinct forms of exon 7), we employ the nomenclature (types "A, B, C, D, E, F, X, Y, Z") suggested by Pauline Yen (Yen *et al.*, 1997) (Fig. 6).

Two features of the first full-length cDNA clone (pDP1678) enabled us to assign it to *DAZ2*. In this cDNA clone, the 5'-most copy of exon 7 (the first 72-nucleotide repeat) is of type "A". In the *DAZ2* genomic locus, the 5'-most copy of exon 7 (within a 2.4-kb genomic repeat) is also type "A." We have found no "A"-type copies of exon 7 anywhere in the *DAZ1*, *DAZ3*, or *DAZ4* genomic loci. Second, the *DAZ2* cDNA clone contained seven tandem "Y"-type copies of exon 7. At the genomic level, each "Y"-type 2.4-kb repeat contains a single *Mlu*I site. An array of appropriately spaced *Mlu*I sites is found in the *DAZ2* genomic locus (Fig. 3). The *DAZ2* cDNA sequence reported here is predicted to encode a 559-amino-acid protein with a molecular weight of 63K. Two previously reported cDNA clones — clone pDP1577 described by Reijo *et al.* (1995), and clone E3 described by Yen *et al.* (1997) — also appear to derive from *DAZ2*.

We assigned the second full-length cDNA clone (pDP1679) to *DAZ4* based on the absence of "Y"-type copies of exon 7. At the genomic level, one or more "Y"-type 2.4-kb repeats are present in *DAZ1*, *DAZ2*, and *DAZ3*, but not in *DAZ4* (Fig. 3). (Note the absence of *Mlu*I sites among the tandemly repeated exon 7's of *DAZ4* [Fig. 3].) The *DAZ4* cDNA sequence reported here is predicted to encode a 487-amino-acid protein with a molecular weight of 55K. Three previously reported cDNA clones — clones E1, E4, and F5 described by Yen *et al.* (1997) — also appear to derive from *DAZ4*.

The third full-length cDNA clone (pDP1680/pDP1681) most likely derives from *DAZ3*, but we cannot exclude the possibility that it derives from *DAZ1*. This cDNA clone differs dramatically from the *DAZ2* and *DAZ4* clones in that it contains a tandem duplication of a 495-nucleotide (165-amino acid) unit. This unit corresponds precisely to exons 2 through 6 and is predicted to encode an entire RRM (RNA recognition motif) domain. The tandem duplication of

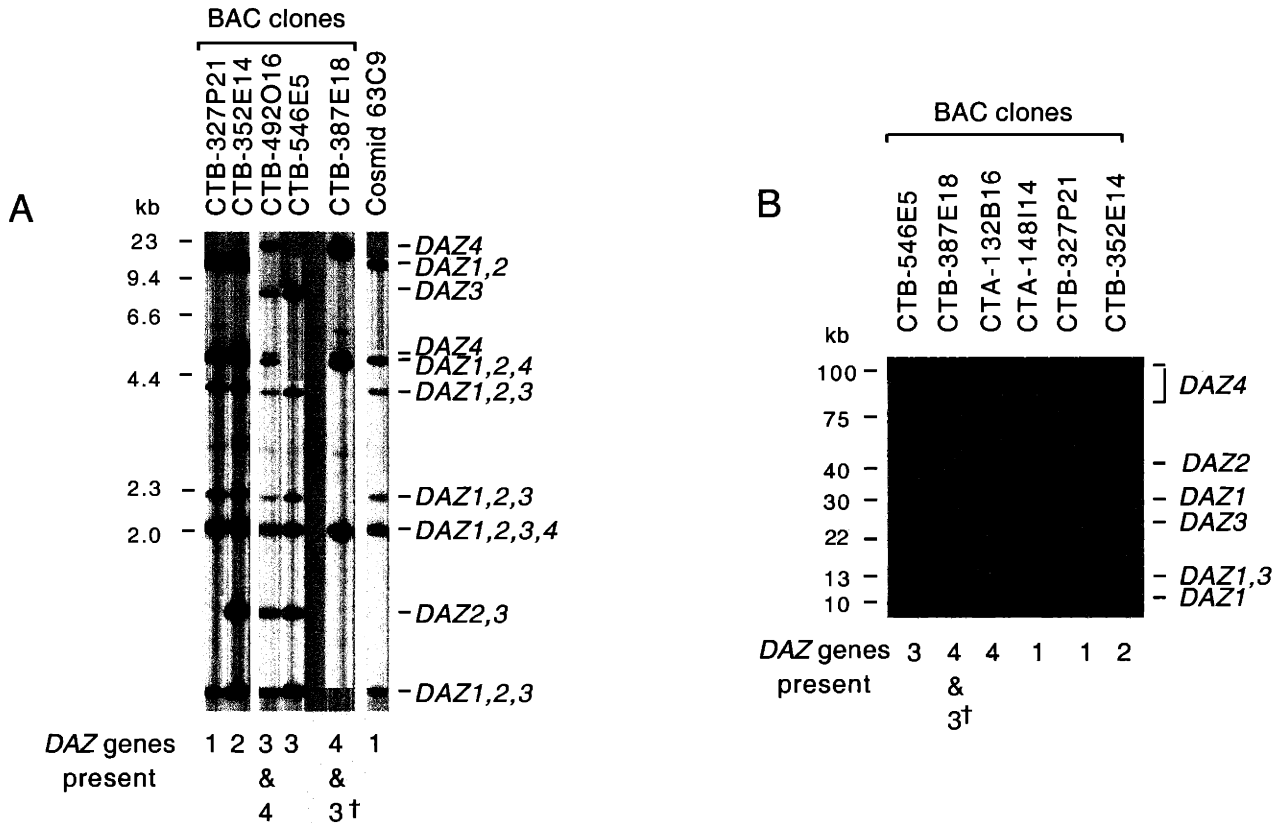


Figure 5a. DAZ 2.4kb repeat probe pDP1649 hybridized to Southern Blot of *Taq*I digested BAC clones containing one or more DAZ genes (listed below the figure) and cosmid 63C9. Note all four genes have bands greater than 2.4kb in length and similar to sequenced cosmid 63C9 which has eight tandem 2.4kb repeats interrupted by a LINE element(Saxena et al. 1996). DAZ genes giving rise to each hybridizing fragment are shown on right. Scale in kb shown at left.

Figure 5b. Southern hybridization of *Mlu*I digested *DAZ* BAC DNA probed with a PCR fragment spanning *DAZ* exons 2 and 3. DAZ genes present in each BAC clone are indicated below. BAC clones CTA-148I14 and CTA-132B16 originate from a different male donor than the other clones.

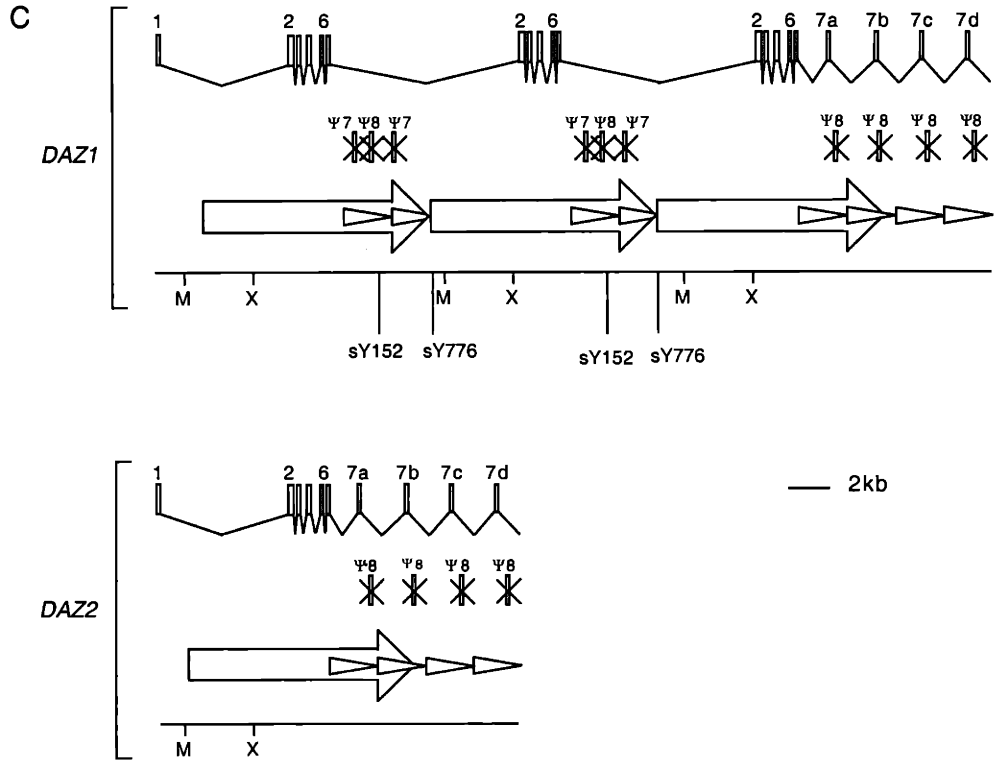


Figure 5c. Schematic diagram of the 5'ends of *DAZ1* and *DAZ2* with three tandem copies or one copy of the 10.8kb repeat which encompasses exons 2-6 and pseudoexons 7, 8 and 7. Exons and pseudoexons are indicated above the repeats, and restriction sites and positions of STS markers sY152 and sY776 (a marker detecting the junction between repeats) are shown below the repeat.

this 495-nucleotide unit within the cDNA corresponds well to the tandem duplication of the 10.8-kb unit in the *DAZ3* genomic locus (Fig. 3). The putative *DAZ3* cDNA sequence reported here is predicted to encode a 579-amino-acid protein with a molecular weight of 65K.

The fourth and fifth cDNA clones sequenced (pDP1575 and pDP1576) were incomplete at their 5' ends, and they most likely derive from *DAZ3*, or possibly from *DAZ1*. Neither cDNA extended sufficiently 5' to include exon 1, but both appeared to derive from transcripts in which exons 2-6 were (at least) duplicated, consistent with their deriving from either *DAZ1* or *DAZ3*. Both cDNAs contained nine copies of exon 7 (BCDEFEXYZ), as also found in putative *DAZ3* cDNA pDP1680, suggesting that these clones may be derived from *DAZ3*.

DISCUSSION

The number of DAZ genes on the Y chromosome

Our FISH analysis and studies of BACs indicate that the human Y chromosome, at least as found in the collection of unrelated individuals studied, contains four *DAZ* genes arranged in two clusters. How can these findings be reconciled with the report of Glaser *et al.* (1998) that there are seven *DAZ* genes or pseudogenes in a single cluster? While it is tempting to attribute these divergent conclusions to polymorphism in *DAZ* copy number — and such polymorphism may exist — this is an unlikely explanation for the discrepancy since both research groups based conclusions on studies of several unrelated men. Instead, we suggest that Glaser and colleagues' data can be re-interpreted in terms of the two-cluster, four-gene model presented here.

As in our studies, Glaser *et al.* (1998) performed FISH on chromatin in differing states of condensation. After hybridizing a *DAZ* cosmid probe to relatively condensed chromatin (from cells in G2 phase or early mitotic prophase), Glaser *et al.* observed two distinct *DAZ* signals (see Fig. 2 in Glaser *et al.*, 1998). These results are in excellent agreement with our two-cluster model.

In high resolution studies on extended chromatin (from cells in G1 or S phase), Glaser *et al.* observed hybridization signals that they interpreted as evidence of seven *DAZ* genes or pseudogenes, some very closely spaced. In these studies, Glaser *et al.* employed probes derived

from *DAZ* 5' regions where, as we now report, some *DAZ* genes carry a tandem duplication or triplication of a 10.8-kb segment. We suspect that repeats within a single gene gave rise to multiple FISH signals that Glaser *et al.,* interpreted as arising from separate genes, leading them to postulate the existence of seven *DAZ* genes, as opposed to our finding of four genes.

Our two-cluster, four-gene model is in good agreement with physical mapping studies conducted by Yen (1997, 1998). Through Southern blot analysis of genomic DNAs, Yen *et al.* determined that most Y chromosomes carry at least three *DAZ* genes (Yen *et al.*, 1997). We suspect that the fourth *DAZ* gene was masked, on Yen's Southern blots, by high nucleotide identity among the *DAZ* genes. More recently, Yen has constructed a YAC-based, long-range restriction map on which she has located three *DAZ* genes (Yen, 1998). Yen identified two *DAZ* genes as forming an inverted pair, in 3'←5'::5'→3' orientation, equivalent to one of the two clusters in our present model. Yen mapped a third *DAZ* gene near an uncloned gap in her YAC contig. We expect that the fourth *DAZ* gene lies within this gap and therefore was not found by Yen.

Sequence family variants

Distinguishing among and unambiguously identifying each of the four *DAZ* genes was technically challenging. We recognized eventually that *DAZ1*, *DAZ2*, *DAZ3*, and *DAZ4* possessed different numbers of intragenic (2.4-kb and 10.8-kb) tandem repeats, but these differences were of little practical use in identifying individual *DAZ* genes, for several reasons. First, both the 2.4-kb and 10.8-kb repeat arrays were far too large to allow PCR amplification across them (as one might do in the case of mini- or micro-satellites). Second, the DNA sequences of the 10.8-kb repeats appear to be identical one to another, obstructing efforts to distinguish among and thereby count the 10.8-kb repeats. Third, many of the BAC clones studied contained portions of two different *DAZ* genes, further confounding gel-based analyses. Apart from these tandem intragenic amplifications, the DNA sequences of the four *DAZ* genes appear to be >99.9% identical (Saxena et al., 1996 and our unpublished results). Consequently, conventional STS-content mapping and restriction fingerprinting of BACs were of little use in distinguishing among the four *DAZ* genes.

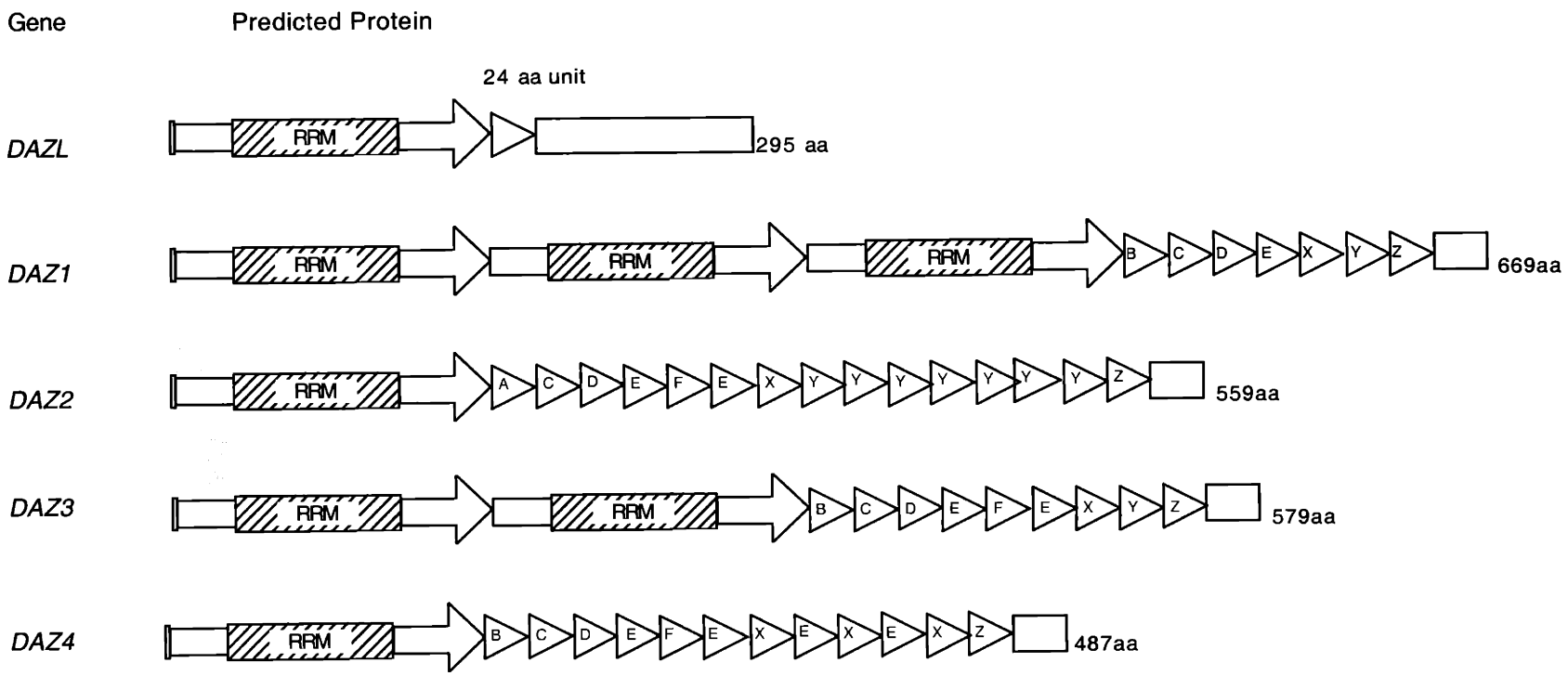


Figure 6 Schematic diagram of predicted DAZ proteins from all human *DAZ* genes. Large arrows indicate the 165aa repeat encompassing the 82 aa RNA binding domain. Arrowheads indicate 24 aa repeats which are labeled according to the nomenclature of Yen et al (1997). The C terminal end of DAZL has no amino acid identity to the C terminal end of DAZ proteins.

In the end, we identified individual *DAZ* genes primarily based on subtle sequence differences — especially base-pair substitutions (Table 1) — that had been revealed by extensive genomic sequencing (Saxena et al., 1996). Since these subtle differences are among members of a gene family on a single Y chromosome, they are not true polymorphisms (which pertain to alleles on homologous chromosomes). We suggest the term "sequence family variants," or "SFVs," to refer to subtle variation (e.g., single nucleotide variation, dinucleotide repeat length variation) between closely related but nonallelic sequences. Based on our experience with the *DAZ* genes, we anticipate that SFVs will play a crucial role in structural and functional analysis of other segments of the human genome that contain families of closely related sequences.

Functionality of the DAZ genes

Based on our genomic and cDNA sequence analysis, we conclude that at least three Y-chromosomal *DAZ* genes — *DAZ2*, *DAZ3*, and *DAZ4* — are transcribed and spliced to encode proteins with one or more RRM (RNA recognition motif) domains. As judged by genomic DNA sequence analysis, the remaining Y-chromosomal *DAZ* gene, *DAZ1*, is also intact, but we have not identified definitively a corresponding cDNA clone. It may prove difficult to identify *DAZ1* cDNA clones — and specifically to distinguish them from *DAZ3* cDNA clones — for several reasons. The *DAZ1* coding region is predicted to be the longest of the four genes (744 aa); it is difficult to capture the entire coding region in a single cDNA clone. This problem is compounded by the likelihood that the 5' portion of the *DAZ1* coding region consists of a perfect tandem triplication of a 495-nucleotide, RRM-encoding unit that is duplicated in *DAZ3*. Finally, the array of exon 7 repeats that is predicted to occur in *DAZ1* transcripts is very similar to that observed in *DAZ3*. Thus, our failure to identify a *DAZ1* cDNA clone should not be taken as evidence that *DAZ1* is a pseudogene. In summary, at least three, and perhaps all four Y-chromosomal *DAZ* genes are translated. Future studies should explore the degree of functional redundancy among the *DAZ* genes and proteins, and specifically whether DAZ proteins with multiple RRMs (as predicted for DAZ1 and DAZ3) differ functionally from DAZ proteins with one RRM (as predicted for DAZ2 and DAZ4).

Reconstructing the evolution of the human DAZ genes

In Fig. 7, we offer a model of *DAZ* gene evolution that combines our present findings with previous observations. We suggested previously that the Y-chromosomal *DAZ* genes arose during primate evolution by i) transposing an autosomal gene (*DAZL*) to the Y chromosome, ii) amplifying and pruning exons within the transposed gene, and iii) amplifying the modified gene (Saxena et al., 1996). The autosome-to-Y transposition apparently occurred in an ancestral Old World primate about 30 to 40 million years ago, after separation from the New World primate lineage (Seboun et al., 1997). Our present findings confirm and refine this model.

We now appreciate that, following transposition of the autosomal gene to the Y chromosome, two different intragenic segments were tandemly amplified (Table 1 and Fig. 7). The first of these tandem amplifications involved a 2.4-kb unit encompassing exons 7 and 8 (Saxena et al., 1996). A subsequent tandem amplification involved a 10.8-kb unit that includes, in addition to exons 2 through 6, one complete and one partial copy of the 2.4-kb unit (Fig. 5c).

The present data also underscore the role of exon pruning during the evolution of the human *DAZ* genes. As recognized previously, most of the 2.4-kb repeats in human *DAZ* genes contain a "pseudoexon," a degenerate, vestigial exon that appears to be excised (as a component of an intron) during processing of *DAZ* transcripts (Saxena et al., 1996). We now conclude, as diagrammed in Fig. 5C, that each of the 10.8-kb repeats in human *DAZ1* and *DAZ3* contains three pseudoexons. Thus, not only the 2.4-kb repeat arrays but also the 10.8-kb repeat arrays appear to be riddled with pseudoexons, at least in humans. In all, the four *DAZ* genes on the human Y chromosome studied here appear to possess a total of 96 exons and 66 pseudoexons. By contrast, their autosomal progenitor, *DAZL*, is a conventionally structured gene with only 11 exons. Remarkably, the reading frames of the Y-chromosomal *DAZ* genes emerged intact from the bouts of intragenic amplification and exon pruning that evidently occurred during evolution. The preserved reading frames suggest that selective pressure on the DAZ proteins was maintained during the evolution of the human *DAZ* genes.

Additional evolutionary insights emerge from comparing the recently described cDNA sequence of Y-chromosomal *DAZ* in *Macaca fascicularis*, a cynomolgus (Old World) monkey (Gromoll et al., 1999) with our present findings in humans. Like some human *DAZ* cDNA clones, the cynomolgus *DAZ* cDNA features a tandem duplication of an RRM-encoding

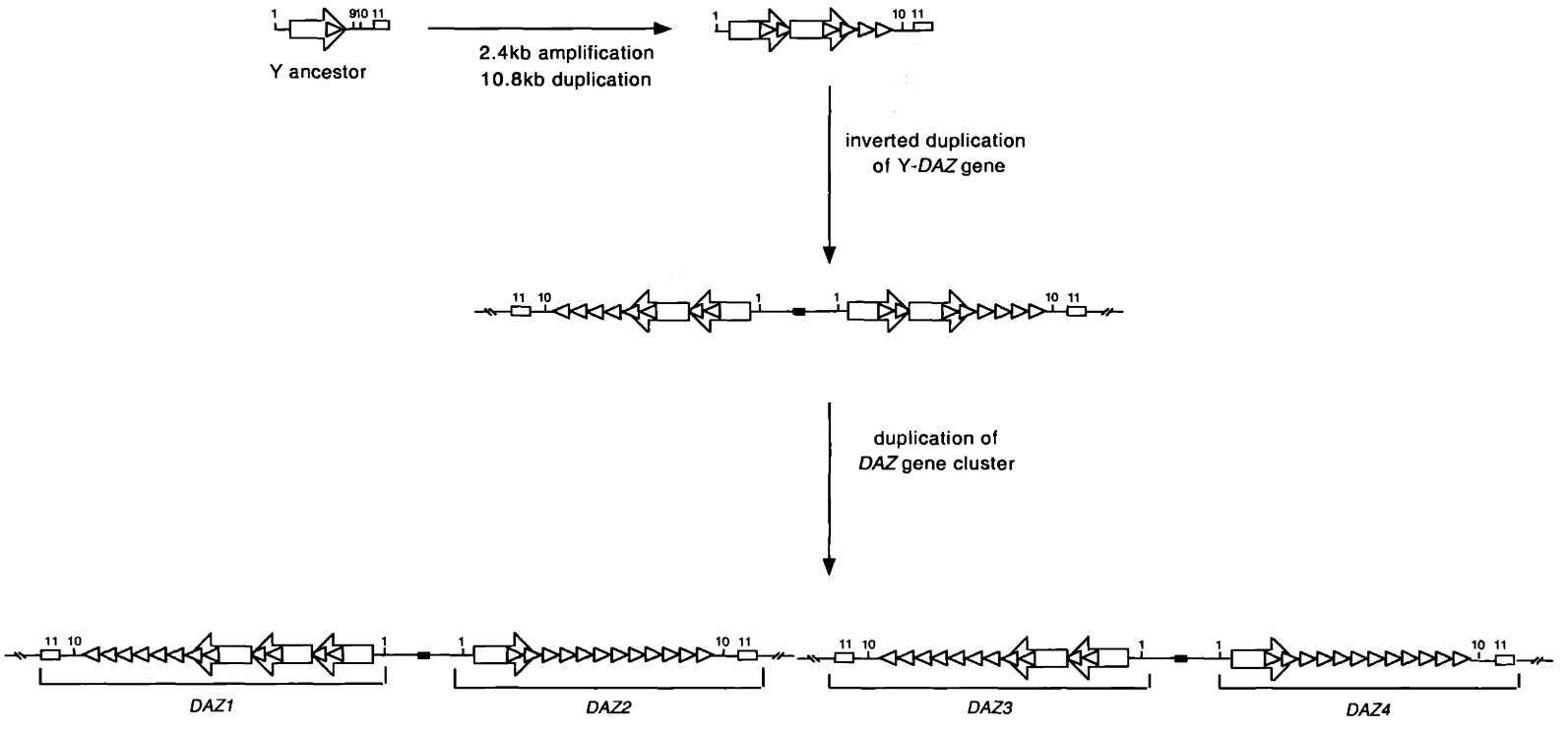


Figure 7. A likely model for evolution of the human Y linked *DAZ* gene family. An ancestral gene transposed to the Y chromosome, accumulated intragenic amplifications, was itself duplicated in an inverted manner, and the entire cluster was duplicated again to result in four *DAZ* genes in modern humans. The 10.8 kb repeat unit in one gene of the inverted duplicated intermediate was lost before duplication of the DAZ gene cluster. The 10.8kb repeat is represented by an arrow, and 2.4 kb repeats are shown as arrowheads. Exons 1,10 and 11 of each gene are shown. A THE element between inverted *DAZ* genes is shown as a black box. Orientation of the DAZ1/ DAZ2 cluster to the DAZ3/ DAZ4 cluster and both clusters to the centromere is unknown.

segment. However, while the duplicated segment in the human cDNAs corresponds to exons 2-3-4-5-6, the duplicated segment in the cynomolgus cDNA corresponds to exons 2-3-4-5-6-7-8-7 (Gromoll et al., 1999). How do we account for this interspecies difference? At the genomic (as opposed to cDNA) level, the 10.8-kb repeat in human *DAZ1* and *DAZ3* contains the following sequence of exons: 2-3-4-5-6-ψ7-ψ8-ψ7 (where ψ denotes a pseudoexon; see Fig. 5C). Taken together, these findings suggest that tandem amplification of the 10.8-kb unit began prior to divergence of the human and cynomolgus lineages — a split estimated to have occurred about 20 million years ago. Further, these results suggest that pruning or degeneration of the last three exons within the 10.8-kb segment began in the human lineage following divergence from the cynomolgus lineage. We conclude that, during evolution, intragenic amplification preceded exon pruning.

ACKNOWLEDGMENTS

We thank the DNA Sequencing Group at the Whitehead/MIT Center for Genome Research for sequencing cosmid 18E8; Isin Dalkilic for assistance in restriction mapping of *DAZ* BACS; Nicolas J. Leschot and Fulco van der Veen for advice and support; and Deborah Berry, Alex Bortvin, Doug Menke, Steve Rozen, Charles Tilford and Tomoko Kawaguchi for comments on the manuscript. Supported in part by NIH.

REFERENCES

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Curr. Biol.* **6**:149-162.

Chen, C., Su, Y., Baybayan, P., Siruno, A., Nagaraja, R., Mezzarella, R., Schlessinger, D., and Chen, E. Y. (1996). Ordered shotgun sequencing of a 135 kb Xq25 YAC containing ANT-2 and four possible genes, including three confirmed by EST matches. *Nucleic Acids Res.* **24**:4034-4041.

Devine, S. E., Chissoe, S. L., Eby, Y., Wilson, R. K., and Boeke, J. D. (1997). A transposon-based strategy for sequencing repetitive DNA in eukaryotic genomes. *Genome Res.* **7**:551-63.

Eberhart, C. G., Maines, J. Z., and Wasserman, S. A. (1996). Meiotic cell cycle requirement for a fly homologue of human *Deleted in Azoospermia*. *Nature* **381**:783-785.

Elliott, D. J., Millar, M. R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T. B., Saunders, P. T. K., Vogt, P. H., Chandley, A. C., and Cooke, H. (1997). Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc. Natl Acad. Sci. USA* **94**:3848-3853.

Fidlerova, H., Senger, G., Kost, M., Sanseau, P., and Sheer, D. (1994). Two simple procedures for releasing chromatin from routinely fixed cells for fluorescence *in situ* hybridization. *Cytogenet. Cell Genet.* **65**:203-205.

Glaser, B., Hierl, T., Taylor, K., Schiebel, K., Zeitler, S., Papadopoullos, K., Rappold, G., and Schempp, W. (1997). High-resolution fluorescence in situ hybridization of human Y-linked genes on released chromatin. *Chromosome Res.* **5**:23-30.

Glaser, 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.

Gromoll, J., Weinbauer, G. F., Skaletsky, H., Schlatt, S., Rocchietti-March, M., Page, D. C., and 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.* **8:**2017-24

Haaf, T., and Ward, D. C. (1995). Higher order nuclear structure in mammalian sperm revealed by in situ hybridization and extended chromatin fibers. *Exp. Cell Res.* **219**:604-611.

Houston, D. W., and King, M. L. (2000). A critical role for Xdazl, a germ plasm-localized RNA, in the differentiation of primordial germ cells in Xenopus. *Development* **127**:447-456.

Kent-First, M., Muallem, A., Shultz, J., Pryor, J., Roberts, K., Nolten, W., Meisner, L., Chandley, A., Gouchy, G., Jorgensen, L., Havighurst, T., and Grosch, J. (1999). Defining regions of the Y-chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. *Mol. Reprod. Dev.* **53**:27-41.

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

Ma, K., Sharkey A., Kirsch, S., Vogt, P., Keil, R., Hargreave, T. B., McBeath, S., and Chandley, A. C. (1992). Towards the molecular localisation of the AZF locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. *Hum.Mol. Genet.* **1**:29-33.

Manz, E., Schnieders, F., Muller Brechlin, A., and Schmidtke, J. (1993). *TSPY*-related sequences represent a microheterogeneous gene family organized as constitutive elements in DYZ5 tandem repeat units on the human Y chromosome. *Genomics* **17**:726-731.

Menke, D. B., Mutter, G. L., and Page, D. C. (1997). Expression of *DAZ*, an azoospermia factor candidate, in human spermatogonia. *Am. J. Hum. Genet.* **60**:237-241.

Ohno, S. (1967). Sex Chromosomes and Sex-linked Genes. Springer-Verlag, Berlin

Redeker, E., Hoovers, J. M., Alders, M., van Moorsel, C. J., Ivens, A. C., Gregory, S., Kalikin, L., Bliek, J., de Galan, L., and van den Bogaard, R. (1994). An integrated physical map of 210 markers assigned to the short arm of. *Genomics* **21**:538-50.

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

Rice, W. R. (1994). Degeneration of a nonrecombining chromosome. *Science* **263**:230-232.

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. *Nat. Genet.* **14**:292-299.

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

Shizuya, H. B., 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.

Silber, S. J. (1989) The relationship of abnormal semen parameters to male fertility. *Hum. Reprod.* **4:**947-953.

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

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

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.

Yen, P. H. (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.

CHAPTER 4

Polymorphism at the Y chromosomal *DAZ* locus in humans

Abstract

The *Deleted in Azoospermia* genes are strong candidates for an *Azoospermia factor* (*AZF*) on the Y chromosome. We recently described the presence of four highly similar *DAZ* genes in humans, arranged in two clusters within the most commonly deleted AZFc region. Portions of a *DAZ* transcription unit, p*49a/*f, were independently identified over fifteen years ago as highly polymorphic Y linked markers, and have been used for population genetic studies extensively since then. These probes detect a complex pattern of over fifteen Y specific bands on *TaqI* Southern blots, of which eight bands are polymorphic. In this study, we determined the molecular basis for the *49a/f TaqI* polymorphism by *DAZ* genomic sequence analysis and restriction analysis of BAC clones containing only one of four *DAZ* genes. The *49a/f TaqI* polymorphism arises by single nucleotide substitution, amplification and deletion of tandem 2.4kb segments within *DAZ* genes, and by duplication and/or deletion of one or the other *DAZ* gene cluster. Analysis of men with known Y chromosomal lineages allows us to begin to infer rates of different types of *DAZ* rearrangements in humans. The *AZFc* region in humans is unstable and is deleted frequently in different Y lineages. Large deletions and/or duplications involving one *DAZ* cluster may have occurred multiple times during human evolution, and tandem repeats within *DAZ* genes are subject to more frequent deletion and duplication. Our results have three important implications- i) as deletions removing two *DAZ* genes may be maintained in some populations, the presence of only two *DAZ* genes, instead of four would appear to be enough for normal male fertility , ii) our understanding of the *49f* polymorphism and other *DAZ* variation now allows the variability at this locus to be exploited more rationally for population diversity and human evolution studies and iii) the DAZ proteins encoded by Y chromosomes from different human backgrounds are themselves highly polymorphic.

Introduction

The Y chromosome has been useful for genetic studies aimed at answering questions about the origins of man, migration patterns of human populations and diversity of current populations (Hammer 1995; Jobling and Tyler-Smith 1995). The Y chromosome allows study of the contribution of men to genetic variability in humans because of its exclusively paternal inheritance and because of lack of recombination enabling Y lineages to be assigned easily. However, from studies of RFLPs and DNA sequence, the Y chromosome is not as polymorphic as the autosomes and the X chromosome(Jobling and Tyler-Smith 1995). Recently, sequence based markers detecting single nucleotide polymorphisms and simple tandem repeats have been exploited to identify population specific Y chromosome haplotypes, allowing the construction of trees tracing evolution of human populations(Hammer et al. 1998; Zerjal et al. 1997).

A notable exception to the low-variability polymorphisms on the Y chromosome is the p*49a*/f *TaqI* polymorphism, which was first described in 1986.(Ngo et al. 1986) p*49a* and *49f* are subclones of Y cosmid 49 (Bishop et al. 1983) and both hybridize to the DYS1 locus on the long arm of the Y chromosome. We showed that DYS1 corresponds to portions of a *DAZ* gene(Saxena et al. 1996). After *TaqI* digestion, probes *49f* and p*49a* detect, in combination, 18 restriction fragments of different sizes, named A through R by decreasing order of size. These fragments are male specific, except for fragments K and L, which correspond to the *DAZ* autosomal homologue *DAZL*. Polymorphic fragments first seen were fragments A C D F I, which were either absent, present or for A and D, variable in size in men from different populations. In subsequent studies, bands B, G and H were also found to be missing in several haplotypes, and additional new bands were occasionally observed. In all, 140 haplotypes have been described, upon analysis of over 1400 men from over 70 different human populations(Poloni et al. 1997). However, researchers working with p*49a*/f *TaqI* haplotypes found them too complex to make direct deductions about relatedness among haplotypes(Poloni et al. 1997). Much population analysis was still conducted using this system, but its variability could not be fully exploited because of its complexity. Apparently identical p*49a*/f *TaqI* haplotypes were seen to arise independently in different Y lineages.

It was interesting that the p49 a/f polymorphism was a result of *DAZ* gene polymorphism because of the association of *DAZ* genes with an important Y chromosomal phenotype. The Y-

linked *DAZ* genes are located within the AZFc region, deletion of which is the most common known molecular cause of spermatogenic failure and male infertility(Reijo et al. 1995). *DAZ* genes, like the three other known genes within this region (CDY, BPY2 and RBM), are expressed in male germ cells and most likely contribute to the phenotype of AZFc deleted men. Of these four gene families, *DAZ* genes have the best credentials as AZFc candidates because homologues identified in other organisms have all proven to be essential only for germ cell development. We recently identified four *DAZ* genes in the AZFc region, and showed that all four appear capable of encoding functional proteins. p*49a/f TaqI* polymorphism suggested that one or more *DAZ* genes was changing rapidly, but the nature of the changes were difficult to interpret because of no clear understanding of the genomic basis of this polymorphism.

In order to understand the molecular nature of *DAZ* polymorphism, we began by deciphering the basis of *49f* polymorphism at the molecular level. We examined if some *DAZ* genes were included within large scale Y chromosomal deletions reported by Jobling (1994(Jobling 1994)). We systematically studied all forms of *DAZ* polymorphism in a sample of men known to have traceably different Y chromosomes based on haplotyping at biallelic and simple tandem repeat loci.

Materials and Methods

Human DNA samples

Men with *49f* long range deletions: Genomic DNA from m38 and m103 was isolated from an agarose plug (gift of M.A. Jobling) for PCR analysis. Genomic DNA from m111 and m119 was provided by M. Jobling.

Men from 30 different Y haplo-groups based on bi-allelic marker typing: Genomic DNA from men already classified into different Y haplo-groups were used for our studies. DNA sources included i) infertile men from the Page lab collection, ii) the Y chromosome consortium (YCC)collection of men with diverse Y chromosomes and iii) the human Polymorphism Discovery Resource.

49f/ TaqI haplotyping

Probe *49f* was prepared by gel purification of the 2.8kb EcoRI insert of plasmid *49f*(J . Weissenbach,1984). Gene-clean purified *49f* probe was radiolabeled by incorporation of alpha

32-P dCTP by primer extension using random hexamers (Boehringer Manheim) or octamers (Gibco-BRL). Taq I restriction enzyme digests of human genomic DNA (5-10 ug) or *DAZ* BAC clones were separated by electrophoresis on 0.8% Seakem agarose gels in 1xTBE for 15-18 hours. DNA was transferred to Genescreen Plus nylon membranes. Filters were baked for 2 h at 80 C. Filters were prehybridized overnight at 42 C in hybridization buffer with formamide and hybridized in the same solution for 48 hours. Hybridized filters were washed at low stringency, twice (15 min each) in 2XSSC at room temperature, and twice (30 min each) in 2XSSC, 0.1% SDS at 42 C. Filters were exposed to film at -80 C for 4 days, and scored according to convention (poloni/ngo).

PCR/ Restriction Digestion of Sequence family variants in DAZ

PCR amplification was performed in 20 ul volumes of 1.5 mM $MgCl_2$, 5 mM NH_4Cl, 10 mM Tris (pH8.3), 50 mM KCl, 100 μM dNTPs, with 1 U Taq DNA polymerase and 1μM of each primer. Primers used were sY581f- CACTGCCCTAATCCTAGCACA, sY581r-TCTTCTGGACATCCACGTCA; sY586f-GTGTGGCACATATGCCTATAAA, sY586r-TTGGTACATCCAGATGCAGAT; sY587f-TGGTTAATAAAGGGAAGGTGTTTT, sY587r-TCTCCAGGACAGGAAAATCC. Thermocycling conditions consisted of an initial denaturation of 3 minutes at 94°C; 35 cycles of 1 minute at 94°C, 1 minute at 64°C,1 minute at 72°C; and, finally, 5 minutes at 72°C. Amplification products were digested with Sau3A (sY581), *TaqI* (sY 586) or DraI (sY587) for 3 hours and run on 4% Nusieve agarose gels.

PCR-SSCP assays

PCR amplification was performed in 10 ul volumes with reagent concentrations as described above with the following modifications: Unlabeled dNTPs were 200uM dATP, dGTP and dTTP, and 2.5uM dCTP, and 0.07 ul a-32P dCTP (3000Ci/mmol) was added to each reaction. 25 cycles of PCR were performed with an annealing temperature of 60 C. 10ul formamide loading buffer (95% formamide, 10mM EDTA, .025% bromophenol blue, .025% xylene cyanol) was added. The samples were denatured at 100 C for 5 mins, cooled on ice, and loaded on a 6% denaturing acrylamide gel, run for 3hrs at 40W. The gel was transferred to Whatman paper and exposed overnight to film at -80 C.

Results

Molecular Basis of *49f/TaqI* polymorphism in humans

49f analysis of *DAZ* BAC clones containing different *DAZ* genes.

We had previously assigned *DAZ* BAC clones to different *DAZ* genes based on sequence family variants and restriction mapping (Chapter 3, Fig.3). We hybridized probe *49f* to Southern blots of *TaqI* digested DNA from these BAC clones and found that each gene has its own characteristic pattern of *TaqI* fragments (Fig 1a), although*49f* is a 2.8 kb fragment containing *DAZ* exons4, 5, 6 and 7a. Analysis of all available *DAZ* sequences, and restriction mapping analysis allowed us to map the *49f* hybridizing fragments back to each *DAZ* gene. The *TaqI* fragments detected by probe *49f* from each *DAZ* gene are shown in Fig 1b-1f. Note that *DAZ4* is polymorphic between the two sources of DNA (CTB, Fig.1e and CTA, Fig.1f)

Analysis of DAZ 2.4kb repeats

Hybridization of the *TaqI* digested BAC filter with a genomic *DAZ* 2.4kb repeat probe (pDP 1649) resulted in a pattern very similar to that seen for *49f* (chapter 3, Fig5a), suggesting that most bands, and all polymorphic bands detected by *49f*, came from the 2.4kb genomic repeats within each *DAZ* gene. This was expected, because probe *49a*, which detects almost all *TaqI* bands detected by *49f* (Ngo et al, 1986), is a 0.8kb fragment that corresponds to intron 7d of *DAZ*. So, in order to understand the nature of polymorphism detected by *49f* and *49a*, we proceeded to study the structure of each genomic 2.4kb repeat. We analyzed available sequence from cosmids 63C9 (*DAZ1*) and 18E8 (*DAZ2*), and from *DAZ1* BAC CTA-148I14 . We found ten forms of 2.4kb repeats in a tandem array(Figure 2a), and each was associated with its own variant form of the 72bp exon 7 (Figure 2b). Pairwise identities between the genomic 2.4 kb repeats, and between the homologous repeat in *DAZL* and all *DAZ* 2.4 kb repeats are shown in Figure 2c. Most repeats are quite diverged from one another, with identities as low as 80%. This likely reflects their ancient origin, as four already diverged exon 7 repeats(corresponding to genomic 2.4kb repeats) were observed in one Y-linked *DAZ* transcript from the macaque

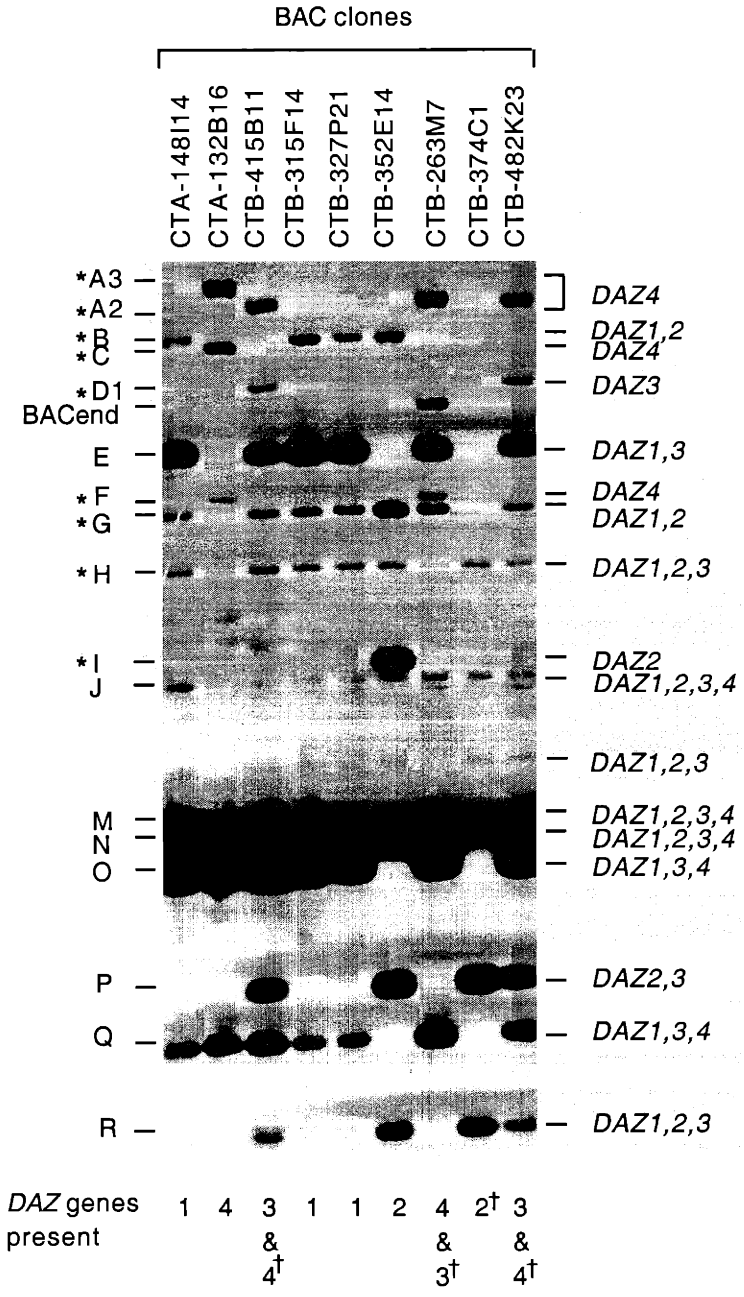


Figure 1a. 49f *Taq*I Southern of DAZ BAC clones (clones described in chapter 3). Hybridizing fragments are labeled A through R by convention (Ngo et al, 1986)on the left of the panel. Polymorphic bands are labeled with an asterisk. On the right are DAZ genes that give rise to each fragment observed. Note that band G comes from DAZ2 in individual CTB, but not in individual CTA. Band R is faint but present in DAZ1 clones.

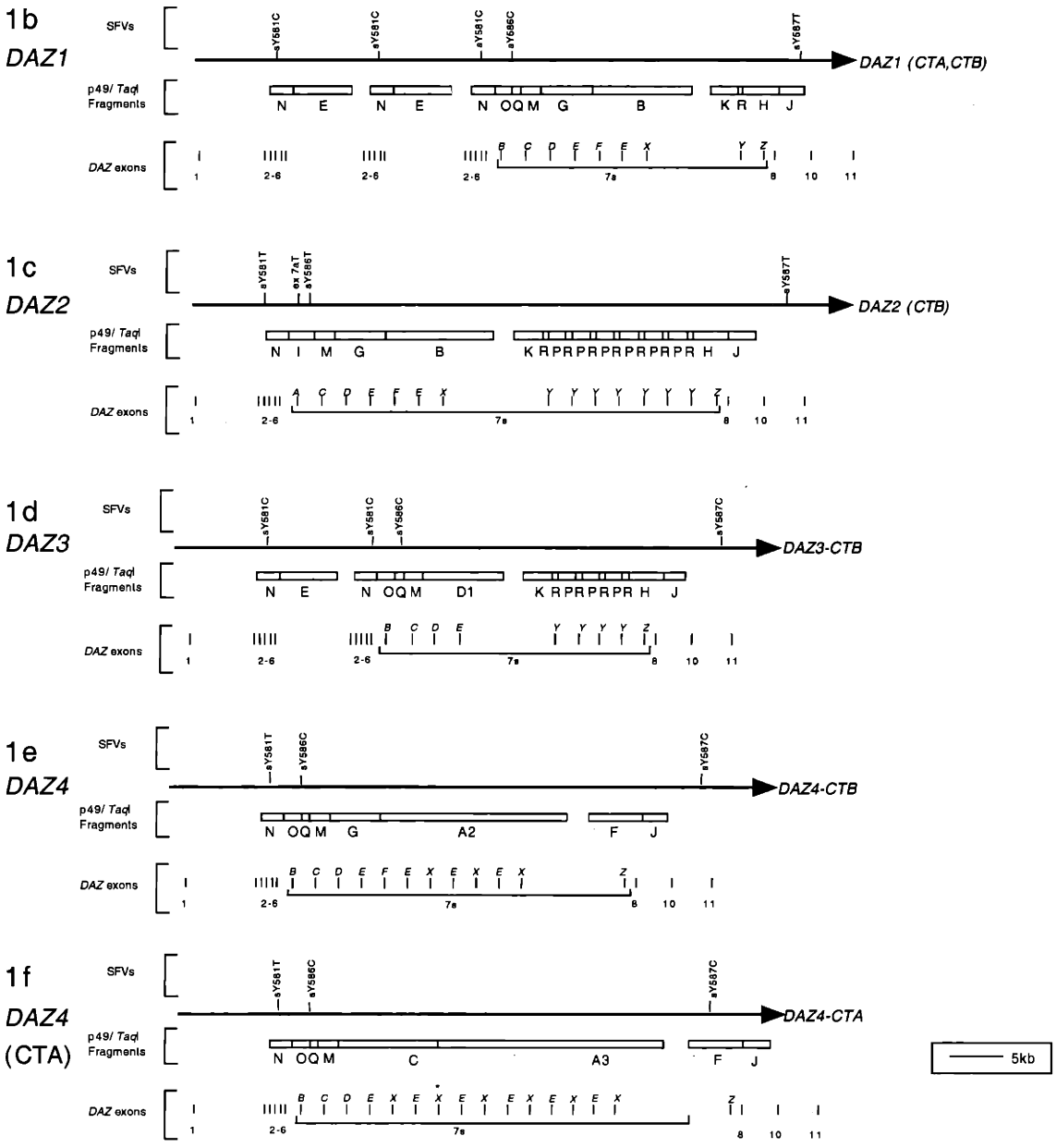
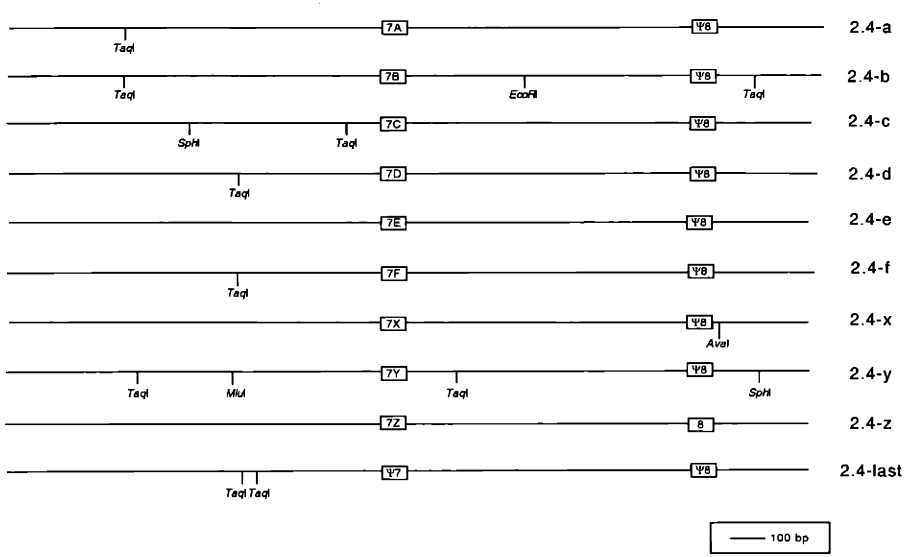


Figure 1b-1f. Schematic diagrams illustrating origin of all 49f/ TaqI fragments within each of four *DAZ* genes. 1b. *DAZ1* in BAC clones from both individuals CTA and CTB, 1c. *DAZ2* in BAC clones from CTB, 1d *DAZ3* in BAC clones from CTB, 1e. *DAZ4* in BAC clones from CTB, 1f. *DAZ4* in BAC clones from CTA. Above the line depicting each *DAZ* gene are sequence family variant nucleotides. TaqI fragments detected by 49f are represented as squares, letters correspond to Ngo et al's classification of fragments (1986). *DAZ* exons are shown below, and exon 7 variants are labeled with the 2.4 repeat type in italics. 2.4kb repeats are labeled according to Yen et al (1997).

2a



2c

	DAZL	A	B	C	D	E	F	X	Y	Z	L
DAZL		86.2	86.0	84.5	85.3	84.4	85.2	84.3	84.7	86.3	88.7
2.4-A			98.9	80.8	82.8	81.7	82.6	81.3	81.6	83.2	84.8
2.4-B				81.6	82.5	81.6	82.4	81.2	81.4	83.0	84.6
2.4-C					89.3	89.7	89.3	87.4	86.8	82.8	82.6
2.4-D						92.8	99.8	92.8	92.9	88.0	83.1
2.4-E							92.7	91.3	91.7	87.2	83.3
2.4-F								92.6	92.7	87.9	82.9
2.4-X									91.1	87.1	87.0
2.4-Y										89.0	82.8
2.4-Z											83.8

2b

```
Consensus  GCATTTCCTG CTTATCCAAG TTCACCATTT CAGGTCACCA CTGGATATCA GTTGCCTGTA TATAATTATC AG
     DAZL  ....A....A ....C....A .......G.. .......T.. .......... .......... .......... ..
     A     ..T.A.T... .........CA .......GG. .......T.. ......G... .....T.... .......... ..
     B     ..T.ACT... .........CA .......GG. .......T.. ......G... .....T.... .......... ..
     C     .A..A....A .......CGA ....G..... .......... .......... .......... .......... ..
     D     C......... .......... A......... .........TG .......... .......... .......... ..
     E     .......... .........A .......... ..A...G... .......... ...C...... ..C....... ..
     F     C......... .......... .......... .........TG .......... .......... .......... ..
     X     .......... .........A .......G.. .......... .......... .......... ..C....... ..
     Y     .......... .......... .......... .......... .......... .......... .......... ..
     Z     .......... .........A ....G..G.. .......... .......... ...C.A.... ..C.....C. ..
  L(Ψ)     A...A....A .........A .......GG. .......... ....G.G... .......... .G........ ..
```

2d

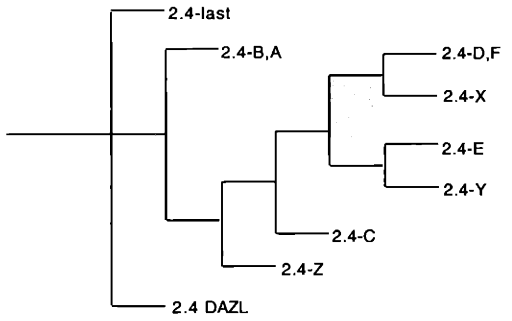


Figure 2. Features of *DAZ* genomic 2.4kb repeats a. Schematic of ten different forms of 2.4 kb repeats, exon and pseudo-exon variant is shown within a box, key restriction sites used for mapping are indicated b. Sequence of the 72 bp exon 7 or pseudo-exon 7 (version L) in *DAZL* and each form of the *DAZ* 2.4 kb repeat. c. Pair-wise comparison of nucleotide identity between genomic 2.4 kb repeats. d. an evolutionary tree of 2.4kb repeats.

(Gromoll et al. 1999). A line drawing illustrating the path of divergence of 2.4kb repeats is shown in Figure 2d.

Apportioning 49f fragments to DAZ genes

Taq I sites within different 2.4kb repeats helped to determine where Taq I fragments mapped within each gene. 2.4kb repeat forms 2.4A and 2.4B are alternative forms of the first 2.4kb repeat in the tandem array within each *DAZ* gene. They are very similar to each other (98.9% identity), but retain reproducible differences, notably, one base pair difference within the exon 7 variant A and B forms, and the presence of a *TaqI* site in 2.4B which is absent in 2.4A. 2.4B therefore, has two internal *TaqI* sites giving rise to a 1.2 kb fragment detected as band Q on the *49f/TaqI* blot (Fig 1a) and present in *DAZ1*, *DAZ3* and *DAZ4* (Fig 1b,d,e,f). Repeats 2.4C, 2.4D and 2.4F have one internal *TaqI* site. Repeats 2.4E, 2.4X and 2.4Z have no *TaqI* sites in them. Repeat 2.4Y has two *TaqI* sites, which are 0.97kb apart, resulting in *49f/TaqI* fragment R. Repeat 2.4L is found at the end of the sequenced genomic clones, has two *TaqI* sites 64bp apart, and the exon 7 variant has not been detected in transcripts thus far.

Different combinations of 2.4kb repeat forms in tandem give rise to particular *49f/TaqI* fragments. The arrangement of 2.4kb repeats in sequenced genomic clones follows a predictable pattern, and this pattern is identical to that observed in exon 7 variants in all *DAZ* transcripts. These are listed in Table 1. The order of exon 7 variants observed in *DAZ* transcripts does not necessarily mean that the genomic locus should contain only these 2.4kb forms in tandem, as *DAZ* transcripts could splice out one or more intervening 2.4 kb repeats. All *DAZ* genes begin with 2.4 kb repeat forms BCD or ACD. Genes with repeats BCD give rise to *TaqI* fragments N (from within exon3 to 2.4B; 2.0kb), O (within 2.4B; 1.9kb), Q (2.4B to 2.4C;1.2kb) and M (2.4C to 2.4D;2.1kb) (figures 1b, d, e). Genes with repeats ACD would give rise to *TaqI* fragments N, I (2.4B to 2.4C;3.1kb), and M. Thus, band I arises from loss of a *TaqI* site in the first 2.4kb repeat.

The types and numbers of the repeats that follow BCD before interruption of the array with a LINE element (in all four *DAZ* genes) determine some *49f/TaqI* fragments. If only 2.4E follows BCD (the pattern BCDE-LINE), then *49f/TaqI* fragment D1 is observed (2.4D to LINE;7.7 kb; *DAZ3*, Fig.1d). Repeats BCDEFEX-LINE give rise to fragment G (2.4D to 2.4F;

Table 1. Order of *DAZ* 2.4kb repeat forms in genomic and transcribed *DAZ* clones.

Sequenced *DAZ* genomic clones

Cosmid 63C9	BCDEFEX-LINE-YZ-last
CTA-148I14	BCDEFEX-LINE-YZ-last
Cosmid 18E8	ACDE-vector

Mapped *DAZ* BAC clones

DAZ1 (CTB)	BCDEFEX-LINE-YZ
DAZ2 (CTB)	ACDEFEX-LINE-YYYYYYYZ
DAZ3 (CTB)	BCDE-LINE-YYYYZ
DAZ4 (CTB)	BCDEFEXEXEX-LINE-Z
DAZ4 (CTA)	BCDEXEXEXEXEX-LINE-Z

		#Transcribed repeats	Reference
DAZ cDNA Clones			
102-*DAZ2*	ACDEFEXYYYYYYYZ	15	Chapter 3
342-*DAZ4*	BCDEFEXEXEXZ	12	Chapter 3
464-*DAZ1,3*	BCDEFEXYZ	9	Chapter 3
466-*DAZ1,3*	BCDEFEXYZ	9	Chapter 3
50-*DAZ3*	BCDEFEXYZ	9	Chapter 3
459-*DAZ2*	ACDEYYYYYYYYZ	14	Reijo , 1995
E3	ACDEFEX--YYYYZ	14	Yen, 1997
E4	BCDEFEXEXEXZ	12	Yen, 1997
E11	ACDEFEXYZ	9	Yen, 1997
DAZ RT-PCR clones			
hT2-A	BCDEYYYZ	9	Yen, 1997
hT2-A'	BCDEFEXYZ	9	Yen, 1997
hT5-A	BCDEYYYYYYZ	11	Yen, 1997
hT5-B	BCDEFEXEXZ	10	Yen, 1997
hT5-C	BCDEFEXYZ	9	Yen, 1997
hT6-A	BCDEXYYYYYZ	11	Yen, 1997
hT6-B	BCDEXEXZ	8	Yen, 1997
hT7-A	BCDEFEXEXZ	10	Yen, 1997
hT7-B	BCDEFEXYZ	9	Yen, 1997
hT8-A	BCDEXEXEXZ	10	Yen, 1997
hT8-B	BCDEFEXYZ	9	Yen, 1997
hT10-A	BCDEYYYYYZ	10	Yen, 1997
hT10-A'	BCDEFEXEXZ	10	Yen, 1997
hT10-B'	ACDEFEXYZ	9	Yen, 1997

4.8kb; Fig1c) and fragment B (2.4F to LINE; 10.1kb; *DAZ1*, *DAZ2*; Fig.1b,c) . Fragment A2 comes from multiple EX repeats following D or F, as these repeats have no internal *TaqI* sites (*DAZ4*; Fig.1e,f). In *DAZ4* from individual CTB, the order is BCDEFEXEX-LINE, with fragment A2, sized 15 kb. (2.4-F to LINE; *DAZ4*, Fig1e). In *DAZ4* of individual CTA, the order is BCDEXEXEX-LINE with fragment A3 sized 20 kb, and an additional fragment C (sized around 9.5kb; Fig 1a) that most likely arises from a point mutation creating a novel *TaqI* site in the second X repeat after D. We have not confirmed that this is indeed the origin of fragment C.

After the LINE element, *TaqI* fragments detected by *49f* probes arise from two combinations of 2.4kb repeats. In *DAZ1*, *DAZ2* and *DAZ3* BACs (Figs 1b-1d), which have one or more 2.4Y type units, fragment K (LINE element to 2.4Y; 2.4kb) and fragment R (within 2.4Y; 0.97 kb) are seen. Two or more copies of 2.4Y in tandem (*DAZ* 2 and *DAZ3*; Figs. 1c,1d) result in *49f/TaqI* band P (2.4Y to 2.4Y; 1.4kb) and a more intense signal for fragment R. *49f/TaqI* band H spans the repeats YZL (2.4Y to 2.4L; 4.0kb). In *DAZ2*, which has no Y type repeat but has only 2.4Z and 2.4L in tandem, fragment F is present (LINE to 2.4L, 5.0kb; Fig1c). Fragment J (3.0kb; 2.4L to intron 8)is the most 3' *TaqI* fragment detected by *49f* probes in all four genes (Figs 1b-1f).

Fragment E (5.9 kb), not within the tandem array of 2.4kb repeats interrupted by the LINE, is seen only in *DAZ* genes 1 and 3 (Fig1a), and comes from 2.4kb repeats within the 10.8 kb repeat (Figs. 1b, d). Band E is not polymorphic; it is present in all individuals tested for *49f/TaqI* polymorphism. An STS detecting the junction between two 10.8kb units, sY776 (Fig 4c, chapter 3), is also present in 80 men with diverse Y chromosomes (data not shown), confirming that men with *DAZ* have at least one *DAZ* gene capable of encoding a protein with multiple RNA binding domains in tandem.

Molecular mechanism of 49f/TaqI polymorphisms in DAZ

Understanding the molecular origin of all *49f/TaqI* fragments in BAC clones allowed us to propose the basis of *49f* polymorphism in men. Our results are summarized in Table 2. Band A, which is variable in length (five variable sizes reported) of which more than one band can co-exist in rare cases, arises from amplification and contraction of the tandem 2.4E and 2.4X repeats. These two repeats amplify and contract as a unit, as evidenced in genomic clones and transcripts observed thus far. Band B, from repeats DEX-LINE or FEX-LINE, is polymorphic in

a few human populations, in whom this band is lost. This band is likely lost in two ways, either by deletion of one *DAZ* cluster (containing genes *DAZ1* and *DAZ2*), or by expansion or contraction of the EX
repeats in both *DAZ1* and *DAZ2*,which would result in bands A or D. Band C, which Spurdle and Jenkins proposed arose in an Caucasoid type chromosome (Spurdle 1992), appears to have arisen as a point mutation creating a *TaqI* site (a rare event) in *DAZ4*, in the fourth 2.4kb repeat after 2.4D (Fig. 1e). Band D arises from contraction of 2.4 kb repeats before the LINE. Band F arises from contraction of 2.4kb repeats after the LINE. Loss of fragment G happens if all DAZ genes with DEF type 2.4kb repeats delete the 4.8kb EF combination. Fragment H can be lost only if all DAZ genes delete 2.4Y repeats after the LINE, or gene conversion occurs from DAZ4 type sequences to the others. Fragment I, as mentioned earlier, is a result of a point mutation in the first tandem 2.4kb unit in any of the four genes.

Large scale deletion and duplication polymorphisms

In 1994, Jobling had reported the presence of a long range SfiI polymorphism detected by the *49f/DAZ* probe which resulted in two continuously variable bands : a large band ranging from720 to 785kb, and a smaller band ranging from 630 to 690 kb (Jobling 1994). In three of 60 world-wide samples that were tested, the smaller fragment was missing, and in two Australian aboriginal samples the larger band was absent. We wondered whether this result suggested that some normal men had only one of the two DAZ clusters. We tested DNA from these men with *DAZ* gene-specific assays (SFV-STSs sY581,sY586 and sY587) and typed microsatellites near *DAZ* . We detected only one of the two *DAZ* clusters in all four cases, consistent with large deletions on the Y chromosomes in these men. 49f/*TaqI* haplotypes of these men are also consistent with the presence of only one DAZ gene cluster (B0 arises from loss of *DAZ1* and *DAZ2*). These results are shown in Table 3. m38 and m119 also lack a 10kb fragment detected by a probe 50f2 that maps close to *DAZ* within the AZFc region (Jobling et al. 1996). As these men have different Y haplotypes, the deletions in all four men are most likely independent events. It is important to note that although these men appear to have deletions, these men may have two DAZ gene clusters which are completely homogeneous with respect to each other.

Table 2. Molecular origin of *DAZ* gene polymorphism detected by *p49f/TaqI* Southerns

***49f/TaqI* bands**	**Allelic forms**	**Molecular Origin**	**Ancestral form**
A series	A1-A5 (14-28kb)	amplification of tandem 2.4EX rpts before LINE	A2 ancestral
	A0	no more than one EX before LINE	
	A3/A2 together	independent amplification of EX in two DAZ genes	
B	B1	2.4DEX before LINE, or 2.4FEX before LINE	B1 ancestral
	B0	i)amplification to 49f/ A and/or deletion to 49f/D in all DAZ genes ii) deletion of one cluster with two B type DAZ genes	
C	C1	may be a point mutation creating novel TaqI site within EX rpts	
	C0		C0 ancestral
D series	D1 or D2	deletion of 2.4 repeats to 2.4DE-LINE or 2.4D- LINE	
	D0		D0 ancestral
F	F1	deletion of 2.4Y repeat after LINE (LINE-Z) in one or more genes	
	F0	i)ancestral form or by ii)deletion of DAZ cluster with F1	FO ancestral
G	G1(4.8kb)	defined by two tandem 2.4 repeats (DEF) before LINE	G1 ancestral
	G0	G is lost by deletion, mediated by 2.4E repeats flanking 2.4F	
H	H1	defined by 2.4Y- 2.4L	G1 ancestral
	H0	H is lost by deletion of 2.4Y after LINE in all four genes	
I	I1	point mutation abolishing TaqI site within 2.4B, or 2.4A	I1 ancestral
	I0	gene conversion, or loss of gene with I1	

Table 3. Large scale deletions/ homogenization involving DAZ genes

Samples	sY 783	sY 758	sY581	sY586	sY587	p49fHaplotype	49f/Sfi I fragment	50f2/c	Y chr haplogroup[a]
m38 (Chinese)	MS	S	T	C	C	(112)A4 B0 C0 D1 F1 I1	755kb	-	VII/VIII
m103 (Caucasian)	L	L	TC	TC	T	(22)A2 B0 C1 D1 F1 I0	740kb	+	IX
m111 (Indian)	MS	S	T	C	C	(39)A3 B0 C0 D0 F1 H0I0	775kb	+	IX
m119 (Australian)	L	L	TC	TC	T	(25)A0 B1 C0 D0 F0 G1 H1 I1	655kb	-	III/IV
OXEN	LMS	LS	TC	TC	TC	(8)A2 B1 C0 D1 F1 G1 H1 I1	650,745	+	VI

[a] Y Haplogroups assigned using biallelic Y markers and shown in figure 3.

Analysis of *DAZ* gene polymorphism in men with known Y haplotypes

In order to estimate the rate of *DAZ* gene variation in human populations, we re-examined single nucleotide polymorphism and *49f* polymorphism in men from 30 different Y chromosomal haplotypes assigned by analysis of non-*DAZ* biallelic markers. These samples serve as indicators of the extent of human diversity, and should allow us to trace occurrence and recurrence of different *DAZ* mutations in different Y lineages.

The results of combined *49f/TaqI* haploytping and analysis of *DAZ* sequence family variants on men from different branches of a known human Y chromosomal lineage tree are presented in figure 3.

AZFc deletions are observed in individuals from multiple different Y haplotypes (S. Rosen and D.C. Page, unpublished results, reproduced here). This is consistent with *AZFc* deletion being a high frequency event that recurs in multiple populations. However, *AZFc* deleted Y chromosomes disappear immediately because of the male infertility phenotype associated with them. It is possible that some Y chromosome haplotypes have higher predispositions to deletion, but the clustering of *AZFc* deletions in some Y haplotypic groups may reflect the bias towards infertile males in the Page Lab collection of cell lines that were used for biallelic haplotyping.

Deletions of one or the other gene cluster may have occurred multiple times during human evolution. PCR assay sY587 detects a single nucleotide variant between two *DAZ* clusters in men from most populations (present in four loci with T in *DAZ1,DAZ2*; C in *DAZ3,DAZ4*). However, in this sample of men, there are several men with only one variant form (C or T). There are two possible means by which this could happen. The first is deletion of a *DAZ* cluster containing one sequence family variant; the second is homogenization by gene conversion of the sequence in the vicinity of one variant form into that of the other cluster. In the first case, only one *DAZ* cluster would remain, and in the second case, two would be present. FISH analysis of *DAZ* in DNA from these men would easily distingish between these two possibilities, but has not been performed in this study. In Y haplotype 92, within haplogroup VIII, all men assayed for sY587 have only the T allele. Men with this haplotype (Tat +, haplotype 92) are deleted for the 50f2/C locus slightly proximal to *DAZ*(Jobling et al. 1996). An intriguing possibility is that these men may have deletions of one *DAZ* cluster and 50f2/c. This Y chromosome appears

stable, as it is observed in all men assayed for this haplotype from a large homogeneous population. This example may provide additional evidence that both *DAZ* clusters are not required for normal fertility. Similarly, men were identified with only sY587C for Y haplogoups III, and VI. These men had fainter hybridization signals for 49f/*TaqI* Southerns (not shown), although equivalent amounts of DNA were loaded in these lanes as for other samples, suggesting they may have fewer copies of *DAZ*. In addition, some 49f/*TaqI* fragments were absent, consistent with one less DAZ cluster. In haplo-group VI, there were single variants sY581T and sY586C associated with sY587C. However, as mentioned above, these variant forms could also have arisen by homogenization.

49f/TaqI Southern analysis shows that different *DAZ* polymorphisms appear to occur at different rates. Change in the copy number of tandem EX repeats has occurred multiple times during human evolution, as seen by the changing size of fragment A. Loss of fragment B is rare. This event could happen either from loss of *DAZ* cluster 1 with two B fragments, or by change in copy number of 2.4kb repeats in both genes *DAZ*1 and *DAZ*2. Fragment C appears stable; once it arose, it was seen in all samples of that lineage, although the sample size here is limited (figure 3). This is consistent with it being a rare mutation creating a novel *TaqI* site. During evolution, *TaqI* sites are more likely to be lost, as the *TaqI* recognition site contains a CpG dinucleotide, which by de-amination of the methylated cytosine into thymine, could result in a C-T transition mutating the *TaqI* site. The *49f* I1 variant is created by the loss of a *TaqI* site, and as shown above, recurs frequently in the human population. It should be noted that there is an opportunity for this to occur in each *DAZ* gene, so although this *TaqI* site was studied in the context of repeat 2.4A, this Taq site could also be lost in 2.4B without any effect on the exon 7 variant. Fragment D was observed in subpopulations of haplogroup IX, and D2 appears to be stable. Absence of fragment F is a rare event, possibly arising from loss of the DAZ cluster containing this fragment. F corresponds to the presence of a Z fragment directly after the LINE element. If the ancestral Y chromosome had this fragment and H, the alternative (LINE-YZ), the only ways to lose either fragment would be by loss of the gene containing them, or truncation of the *DAZ* gene with these fragments. G appears to be present in multiple *DAZ* genes, and is also rarely absent.

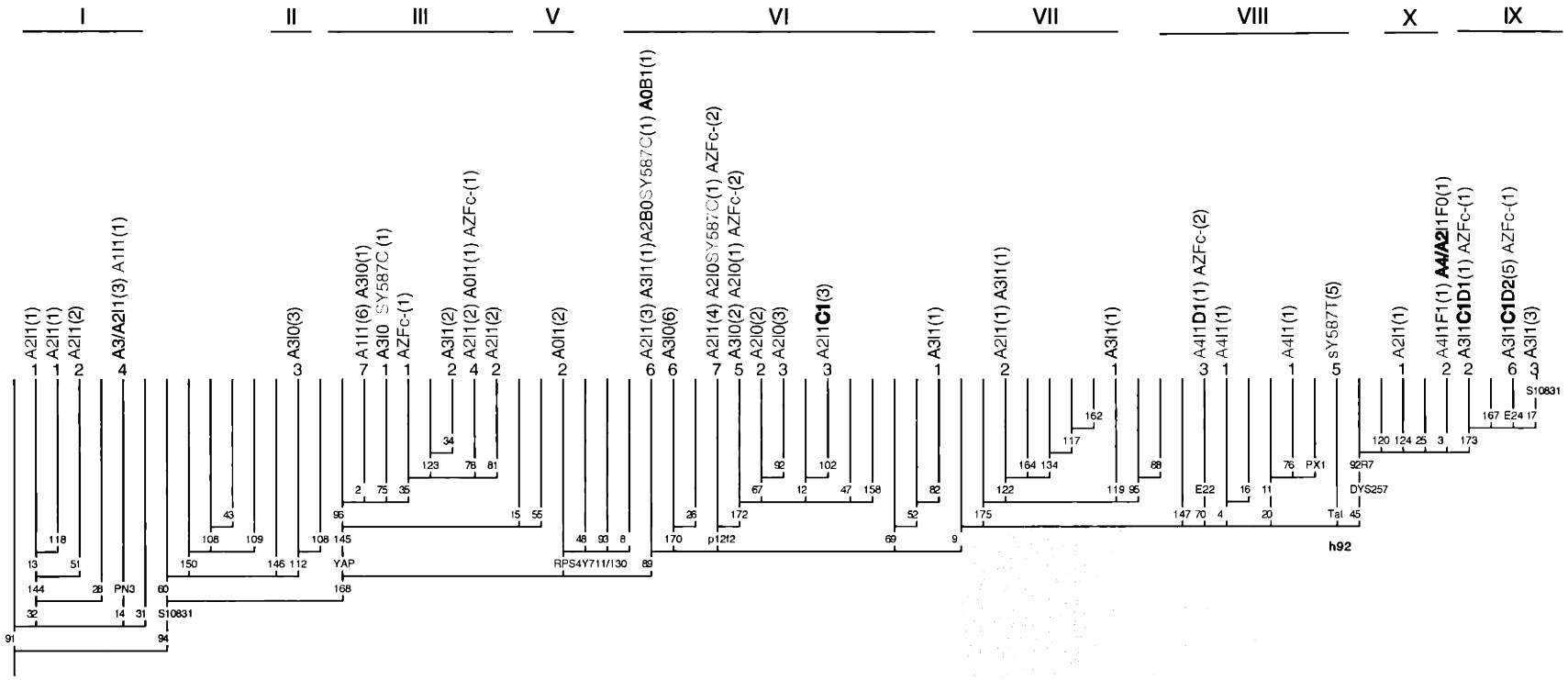


Figure 3. DAZ polymorphism analysis (49f/ TaqI haplotypes and results from SFV-STSs) of men from known Y chromosomal lineages. Schematic diagram of a Y chromosomal tree based on biallelic markers; above the tree are numbers indicating the number of men haplotyped for DAZ polymorphism, and above are polymorphic 49f/ TaqI fragments observed. Above, outlined in green, are results from SFV-STS marker sY587, only shown if one of two variant forms was detected. In brackets are the number of individuals with each haplotype. Samples for AZFc deleted men are labeled. Roman numerals indicate the haplogroup of each set of individuals. h92 below the branch with Tat45 is haplotype 92. (See text). Scoring of I1 band was 100% concordant with the observation of an sY586 T variant.

Discussion

DAZ genes differ by amplification and deletion of internal genomic segments

Our analysis of *DAZ* BAC clones shows that, in the individuals from whom these clones were generated, *DAZ* genes differ from each other mostly in the extent of amplification and contraction of different internal repeats. The major features that distinguish *DAZ2* and *DAZ4* from *DAZ1* and *DAZ3* is contraction of the 10.8kb unit. A protein with one RRM is encoded by *DAZ2* and *DAZ4*, instead of two or even three RRM domains predicted in proteins from *DAZ3* and *DAZ1*. *DAZ* transcripts from the rhesus macaque (a lineage which separated from humans over 20 million years ago) appear to encode two RRM domains. *DAZ* genes also differ from each other in the numbers and orders of internal 2.4kb repeats which encode exon 7 in almost all repeats, and exon 8 in one repeat. *DAZ* proteins therefore differ from each other in the number of RNA binding domains and the number and type of 24 aa repeats.

Nature of DAZ gene polymorphism in humans

Our analysis of *DAZ* variation in men with different Y lineages , although not comprehensive, suggests that *DAZ* genes have changed considerably in human populations. Some evidence points to the stable propagation of Y chromosomes with only one *DAZ* cluster, but this can only be said with confidence if this is confirmed by FISH. The Y chromosome is known for its lack of variation, as mentioned earlier, and therefore it may be premature to presume that perfect sequence identity in and around *DAZ* would certainly mean a deletion of one *DAZ* gene cluster. All the data presented here is consistent with extreme homogenization within the *AZFc* region instead of deletions of one DAZ cluster.

In addition to the *49f/TaqI* polymorphism(which detects single nucleotide variation, variation in copy number of 2.4kb repeats, and deletion or homogenization of one cluster of *DAZ*) we believe even more variation exists within *DAZ*. Amplification and deletion of the 10.8 kb repeat is already apparent within *DAZ* genes in one individual, and it is highly likely that copy numbers can vary between individuals. Some experimental evidence already exists for this; new

DAZ1 clones from a different individual have only a duplication of the 10.8kb unit instead of a triplication (H. Skaletsky, D.C.Page,Y chromosome sequencing project). Another source of variation are the 2.4Y repeats; they form long perfect tandem arrays, and probably expand or contract in numbers at a more rapid pace than other 2.4kb repeats. This tandem repeat is not detected by the *49f/TaqI* Southern digest, as 2.4Y has two internal *TaqI* sites. Spurdle and Jenkins looked at *49f* polymorphism using other enzymes, and found band-patterns that did not correlate with *TaqI* haplotypes. (Spurdle and Jenkins 1992) These may have included variation in 2.4Y arrays. Other forms of polymorphism not described here, but equally likely are duplications of entire clusters and inversions. In fact, by FISH, there is a large Y chromosomal inversion in some Gujarati Indian men that results in one DAZ cluster on the short arm of the Y chromosome (J. Hoovers, pers. comm.).

DAZ gene polymorphism and DAZ protein polymorphism

Although the Y chromosomal 49f/*TaqI* polymorphism has been known for about 15 years, it was never anticipated that this polymorphism was actually in a functional gene family implicated in a key male specific function - spermatogenesis. For population studies, non-coding loci are preferred, as they are selectively neutral and accumulate unrestricted mutations. Y chromosomal variants are few, yet the *DAZ* genes stand out as being so highly polymorphic both at the genomic and protein level. DAZ proteins differ in individuals with different Y chromosomal haplotypes, as the 24 aa units X, Y and Z each differ from that of DAZL by three amino acids, units A, B, C, and E differ from that of DAZL by five amino acids, and units D and F differ from DAZL's by seven amino acids. Different tandem combinations and numbers of different 24aa repeats would result in rather polymorphic DAZ proteins. This may have some or absolutely no functional consequence, but in any case is the first case of a Y-encoded protein with so much variation.

Implications on Male Infertility

Our original aims for these studies were to identify if any of the four DAZ genes were specifically required for normal spermatogenesis (and so would normally be unchanged in populations), or whether a subset of specific genes could be lost without any detriment. Our

examination here has instead shown that Y linked *DAZ* genes are quite fluid, and differ from each other only in the rearrangement events different genes have experienced in each population. In the individuals from whom BAC clones were derived, *DAZ2* and *DAZ4* have lost the10.8 kb duplication originally seen in macaques, whereas *DAZ1* now has a triplication. In addition, one or the other cluster may be independently deleted in different Y chromosomal lineages, and was likely maintained in at least one such lineage, suggesting that only two genes may be needed for normal male fertility. This would mean that *DAZ* genes from either of the two clusters produce enough protein for normal spermatogenesis. However, it is also possible that men with deletions of one DAZ cluster are sub-fertile, and these Y chromosomes are slowly being eliminated from the lineage. This would be consistent with the observation of rare sY 587C individuals in multiple populations. If that is so, men with the Tat + allele, in haplotype 92, may have developed a compensatory mutation to maintain the chromosome in that population. The correlation between different DAZ protein variants and the efficiency of spermatogenesis in men with different Y haplotypes should perhaps be examined in detail.

Studies of DAZ evolution in humans

Based on our description of the molecular basis of *49f* polymorphism, the conclusions drawn about the origins and migrations of human populations by typing 4000 men in the past with *49f* can be re-examined, and understood much more clearly. PCR assays can be developed to assay some of the variations detected by *49f TaqI* southerns, as has already been done in the case of 49f/*TaqI* fragment I; this point mutation is detected by SFV-STS assay sY586. The data available from earlier studies might also be used for a deeper examination of DAZ mutation rates as described below.

From the studies presented here, we determined which *DAZ* polymorphic variants occur infrequently, and which are more common. We looked for specific *DAZ* mutations that persist, which might suggest some selective advantages to those mutations, but our analysis did not give a strong indication of such mutations. We also looked for mutations that are common but do not persist, possibly because of detrimental effects to male fertility, and saw that AZFc deletions are such events(as expected), and some deletions/homogenizations of one cluster (may also fall into

this category. An analysis of more diverse Y chromosomes using this and other assays of DAZ polymorphism may be able to uncover selectively advantageous or detrimental DAZ mutations, if any exist..

Acknowledgments

I would like to thank Steve Rosen for generously providing information about the Y chromosomal tree based on biallelic haplotyping he performed, and for his comments. I am also grateful to Mandy Spurdle for 49a haplotypes of some African individuals and Mark Jobling for gifts of DNA samples from men with SfiI long range polymorphisms.

References

Bishop CE, Guellaen G, Geldwerth D, Voss R, Fellous M, Weissenbach J (1983) Single-copy DNA sequences specific for the human Y chromosome. *Nature* 303:831-832.

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. *Human Molecular Genetics* 8:2017-24.

Hammer MF (1995) A recent common ancestry for human Y chromosomes. *Nature* 378:376-378.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, et al (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 15:427-41.

Jobling MA (1994) A survey of long range DNA polymorphisms on the human Y chromosome. *Human Molecular Genetics* 3:107-114.

Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, et al (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Human Molecular Genetics* 5:1767-1775.

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449-56.

Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA Probe Detecting Multiple Haplotypes of the Human Y Chromosome. *American Journal of Human Genetics* 38:407-418.

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti A, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Americal Journal of Human Genetics* 61:1015-35.

Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, et al (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genetics* 10:383-393.

Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, et al (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.

Spurdle A, Jenkins T (1992) Y chromosome probe 49a detects complex PvuII haplotypoes and many new TaqI haplotypes in southern African populations. *American Journal of Medical Genetics* 50:107-125.

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, et al (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 60:1174-83.

CHAPTER 5

Conclusions and Future Directions

The work presented in this thesis has been a detailed study of the human *Deleted in Azoospermia* gene family which, like homologues in model organisms, most likely plays an important role in germ cell development. In humans, the *DAZ* genes are implicated in male infertility as they lie within the *AZFc* region of the Y chromosome, deletion of which causes severe spermatogenic failure and male infertility. My work focused on characterizing all members of the gene family in humans, assessing variation in *DAZ* gene content in fertile men, and the beginning of examination of *DAZ* gene function by analysis of *AZFc* region deletions in infertile men and by complementation of meiotic defects in *Drosophila*. In addition, this thesis provided some more general insights into Y chromosomal genomic structure and Y chromosomal evolution.

Description of the human *DAZ* gene family

When *DAZ* was identified as the first *AZF* candidate gene in 1995, it was thought to be a single copy, testis specific gene with no known human homologues. My work revealed the existence of a highly similar human homologue, *DAZL* (formerly called *DAZH*), that is not within the *AZFc* region, or even on the Y chromosome. *DAZL* too is germ cell specific in both males and females. I further found that there are multiple Y linked *DAZ* genes that map entirely within the *AZFc* region, and these genes are highly modified derivatives of an ancestral gene. The open reading frames of *DAZ* genes are maintained even after many amplification events, and after systematic splicing out of different 'pseudo-exons', suggesting that *DAZ* genes are still functional. An in-depth understanding of the genomic structure of each *DAZ* gene and analysis of transcripts originating from different genes suggest that all four Y-linked *DAZ* genes are most likely functional, and encode slightly different protein variants. There are two internally repeated segments within some *DAZ* genes, and one results in transcripts with potential to encode *DAZ* proteins with multiple RNA-binding domains. This may be a functionally significant difference between *DAZ* genes, as multiple RRM domains may change the nature or affinity of interaction of *DAZ* with its target(s).

Analysis of *DAZ* polymorphism in men reveals that *DAZ* proteins are not only different within an individual, but are different between individuals as well. Tandem amplification or

deletion of internal genomic segments --which allows different Y-linked *DAZ* family members to be distinguished from one another in one individual-- is also an ongoing process in all *DAZ* genes on the Y chromosomes of men from diverse populations. Therefore, *DAZ* proteins have many variant forms, and each individual has a subset of these protein forms. The functional consequences of this more subtle variation will require further investigation. The picture that emerges from this study is that the germ cell-specific *DAZ* family is rapidly evolving, even within the human lineage. As each Y-*DAZ* gene is not stable in the population, it is impossible to know if any gene is more critical within the *AZFc* region than any other.

The multi-copy nature of this gene family raises questions about overlapping functions of DAZ and DAZL proteins. Is the *AZFc* deletion phenotype variable because of the partial compensation by *DAZL* for *DAZ* function? *DAZL* could be a non-*AZFc* region modifier in *AZFc* deleted men. Is there partitioning of germ cell functions between the different proteins? These questions are as yet unanswered, as in-depth studies of the expression and localization of both proteins during all stages of germ cell development have not been carried out. The RRM domains of DAZ and DAZL proteins are very similar, but amplifications within DAZ and deletion of the C terminal domain may change the properties of DAZ proteins. *DAZ* homologues in the model organisms mouse, frog, fly, fish and worm are all orthologues of *DAZL* and not *DAZ*.

My studies of this gene family in humans open many avenues for future research. Studies to define the role of *DAZ* proteins in humans will need combined genetic and biochemical approaches. I have initiated work in two paths that merit further investigation -- a) studies of *AZF* region deletions in infertile men to understand the contribution of *DAZ* to the *AZFc* phenotype, and b) complementation of mutant phenotypes in model organisms using human *DAZ* proteins. Another key future direction is more biochemical in nature - to identify *DAZ* and *DAZL* targets and interacting proteins. This may lead to the identification of more germ cell-specific components, as *DAZ* family proteins may be part of fundamental, and conserved germ cell processes. Evolution of DAZ into a multi-gene family in primates may have added novel germ cell functions to the repertoire already existing in lower organisms. Furthermore, molecular biological studies of *DAZ* and *DAZL* promoters may reveal some exciting clues about determinants of germ cell specificity, an area of investigation that has remained unexplored.

DAZ and *AZFc*-deleted men

The analysis of atypical deletions in the *AZFc* region presented in this thesis is an important first step towards understanding the *AZFc* region deletions and their correlation with the variable *AZFc* phenotype. This is the first report of mapping a breakpoint in an *AZFc* deleted patient. It is unexpected that the deletion is mediated by a genome typical repeat, an HERV endogenous retroviral element, in a region so full of Y specific repeats. Mapping of other *AZFc* region breakpoints will be important in identifying what features in this region are hot-spots that promote recombination events. Analysis of *DAZ* gene copy number polymorphism, the high frequencies of *AZFc* region deletions, and these atypical *AZFc* region deletions all point to the instability of the genomic territory in this region of the Y chromosome. My observations also suggest that *AZFb* region deletions may encompass deletions of one *DAZ* cluster, again raising the possibility that deletions of different gene combinations, and reduction in copy number of some genes all contribute to different extents of spermatogenic failure.

The phenotype of *AZFc* deleted men most likely results from a very complex combination of genetic and possibly environmental factors. The role of *DAZ* family proteins in causing this heterogeneous phenotype will be difficult to determine. It is already clear that *DAZ* proteins in other organisms can function during different stages of germ cell development, and absence of *DAZ* (or lowered doses of DAZ family proteins) at any point during germ cell development could be responsible, at least partially, for the phenotype seen in these men. The *AZFc* phenotype appears to become more severe with age, and the simplest explanation for the variable, progressive phenotype is that *DAZ* family proteins, with other *AZFc* region spermatogenic factors, reduce spermatogenic output by disrupting spermatogonial stem cells, either by affecting survival, or differentiation. Multiple effects at earlier stages -- during germ cell migration to the gonads, or during proliferation -- and at later stages -- in meiosis and in spermiogenesis -- are also possible points of action of *DAZ* genes. Combined knowledge from the phenotypes of men with various *AZFc* region deletions, and about the many roles of DAZ family proteins in germ cell development in different organisms will the role of this versatile gene family in human infertility be understood.

My work has underscored the genomic complexity of the Y chromosomal *AZFc* region and has been an important first step in determining the underlying causes of infertility of *AZFc*

deleted males. Definition of a model for the arrangement of the four *DAZ* genes in humans has been an important step in understanding the structure and function of the *AZFc* region. The model aided the scaffolding of sequenced clones in the *AZFc* region. STSs used to distinguish between sequence family variants, first used for *DAZ* clones, have been critical in assigning BAC clones to different copies of repeated segments to aid sequencing efforts of the *AZFc* region and the entire Y chromosome. Now that we can assume that all *DAZ* genes are functional, FISH analysis using *DAZ* probes can be used as a tool to study changes in *DAZ* structure in infertile and fertile men.

Study of human DAZ proteins in model organisms

Rescue of the *boule* mutant phenotype by human DAZ proteins, although preliminary, illustrates that human and mouse DAZ proteins are capable of compensating for boule in meiosis and are functionally conserved with boule to some extent. This is the first demonstration that mammalian *DAZ* proteins can function in meiosis and indeed that a human Y encoded *DAZ* protein is capable of any function at all. These results provide an exciting possibility for further investigation - a more rigorous analysis of the extent to which *DAZ* proteins can function in meiosis. Ectopic expression of human *DAZ* and *DAZL* proteins under the control of the boule promoter, in order to recapitulate boule expression, could now be performed. Furthermore, all four human Y-linked *DAZ* genes could be expressed in flies and subtle differences between efficacy of the proteins to function in meiosis could be examined.

Similar analyses can be performed in other model systems in the future to study the functional conservation of *DAZ* proteins across the species, and to study the sex-specific roles of *DAZ* in germ cell development. Over-expression of human Y-linked *DAZ* genes in the mouse and rescue experiments in the mouse based on knowledge of the genomic organization would also allow a deeper understanding of the role of *DAZ* proteins in female germ cell development.

Insights from *DAZ* Evolution

The study of *DAZ* genes has provided a fascinating example of the evolutionary paths of sex chromosomes across millions of years during primate evolution and during evolution within

the human species. Our report of *DAZ* transposition to the Y in primates was the first demonstrated example of the transposition of a fertility gene from an autosome onto the differentiated Y chromosome without an X intermediate. This evidence suggests that there might be a selective force that promotes accumulation of fertility factors on the Y chromosome. Sequestration on the Y would be an especially convenient way to prevent expression in the female - if fertility factors were detrimental to females. However, regulation of gene expression such that female detriment genes are shut off in females is also not a difficult option. In the case of *DAZ* genes, there may be no female detriment,. Instead, mammalian DAZ family proteins seem to be required during early germ cell development in both sexes.

Discovery of multiple testis-specific families on the Y chromosome with no known X homologues has suggested that, as we proposed, much of the chromosome may have acquired fertility factors from autosomes. An example of autosome-to-Y transposition of a fertility factor has also been described in *Drosophila* (Kalmykova et al. 1997). Lahn and Page described the retroposition of another testis-specific gene in the *AZFc* region, CDY, suggesting another mechanism of accumulation of spermatogenic factors on the Y(Lahn and Page 1999). More examples may become apparent as more genes are found from genomic efforts on Y chromosomes from other species. The selective force leading to accumulation of spermatogenesis genes on the Y chromosome may involve random transpositions and retention of a subset of genes that are particularly advantageous to male fertility.

After transposition to the Y chromosome, amplifications within *DAZ* and of entire *DAZ* genes were events caused by the unique genomic structure of the Y chromosome, and perhaps because of proximity to subtelomeric regions. Amplification of *DAZ* genes might have been a means to compensate for the inevitable degeneration of *DAZ* genes, like all genes on the Y chromosome. Within humans too, *DAZ* genes are changing because of the many internal repeats, and possibly because these genes are within an unstable region close to the heterochromatin. Although *AZFc* region deletions removing *DAZ* arise frequently, they are selected out of the population, and so, *DAZ* genes are maintained intact in human Y lineages.

In conclusion, this thesis work provides a solid foundation for future research on *DAZ* gene function in humans --- the genomic structure and steps in evolution of the repetitive Y chromosomal *DAZ* genes have been well characterized here, and the unexpectedly high variability in humans has been explained here. This work lays the foundation for further

investigation into the role of human DAZ in male fertility, and takes small steps in that direction by examination of AZFc region breakpoints, and complementation studies using a human protein in flies.

References

Kalmykova AI, Shevelyov YY, Dobritsa AA, Gvozdev VA (1997) Acquisition and amplification of a testis-expressed autosomal gene, SSL, by the Drosophila Y chromosome. *Proc Natl Acad Sci U S A* **94**:6297-302.

Lahn BT, Page DC (1999) Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nature Genetics* **21**:4290433.

Appendix 1

Characterization of atypical *AZFc* region deletions in infertile men

Introduction

AZFc deleted men exhibit a wide range of phenotypes, all of which fall within the clinical definition of male infertility(Reijo et al. 1995). Sperm counts of *AZFc* deleted men range from zero to 0.5% (1 million /ml) of normal(Reijo et al. 1996) . Testis biopsies from some *AZFc* deleted men suggest an even more complicated picture; men with azoospermia have variable testicular histologies(Reijo et al. 1995; Vogt et al. 1996). Some men have no germ cells in seminiferous tubules of their testes (Sertoli cell only phenotype), some men have an excess of spermatocytes arrested at meiosis I and no signs of post-meiotic cell types (maturation arrest phenotype), and some men have various combinations of the above, with seminiferous tubules containing no germ cells, and tubules with arrested spermatocytes. These men may even have some tubules with small foci of normal spermatogenesis, and may produce a few normal sperm which are capable of fertilizing an oocyte using assisted reproduction techniques(Mulhall et al. 1997; Silber et al. 1998). Testis biopsies are not routinely performed on oligospermic men, therefore histological phenotypes are not available for these men. This variation in extent of progression of spermatogenesis in different *AZFc* deleted men suggests i) that genes within the *AZFc* region are not absolutely required to make normal sperm, but are instead required for normal levels of sperm production and ii) the phenotypes of *AZFc* deleted men vary because of as yet uncharacterized factors, which may include micro-environmental influences, and non *AZF* genetic backgrounds.

It is possible that *AZFc* deletions in men have very different breakpoints, and phenotypes vary depending on combinations of genes deleted or on the number of copies of different genes deleted. Some evidence already exists for a correlation between severity of spermatogenic failure and extent of *AZFc* region deletion. In 1998, Silber et al reported a study of 19 men with Y deletions encompassing the *AZFc* region(Silber et al. 1998). They found that all five men with Y deletions larger than the *AZFc* region (intervals 6D-6F) had absolutely no sperm in testis biopsies, whereas of 14 men with deletions limited to intervals 6D-6F, all but two (86%) had sperm in testis biopsies. These findings suggest that deletions of additional genes in men with larger *AZFc* deletions cause a more severe testicular phenotype.

A factor complicating the precise determination of deletion breakpoints in infertile men is the repeat rich nature of the *AZFc* region. The physical organization of the *AZFc* region is at last

beginning to be understood, as sequencing of this region is coming to a close (Y chromosome sequencing project, H. Skaletsky and D.C. Page, unpublished observations). The *AZFc* region is at least 3.5 Mb in length, and does not contain any single copy sequence. STS analysis suggests that 3/4 of all *AZFc* deletions have breakpoints that lie within 250kb of flanking sequence with over 99.9% identity. Markers that distinguish between different copies of all genes and repeats specific to this region (SFV-STSs) have enabled us to begin examining Y chromosomal breakpoints of atypical *AZFc* deleted patients. In this appendix, I will briefly discuss our progress in defining the breakpoints of three interesting patients, and comment on the gene content lost from their Y chromosomes.

Methods

PCR-SSCP assays

PCR amplification was performed in 10 ul volumes with reagent concentrations as described above with the following modifications: Unlabeled dNTPs were 200uM dATP, dGTP and dTTP, and 2.5uM dCTP, and 0.07 ul a-32P dCTP (3000Ci/mmol) was added to each reaction. 25 cycles of PCR were performed with an annealing temperature of 60 C. 10ul formamide loading buffer (95% formamide, 10mM EDTA, .025% bromophenol blue, .025% xylene cyanol) was added. The samples were denatured at 100 C for 5 mins, cooled on ice, and loaded on a 6% denaturing acrylamide gel, run for 3hrs at 40W. The gel was transferred to Whatman paper and exposed overnight to film at -80 C.

PCR/ Restriction Digestion of Sequence family variants in DAZ

PCR amplification was performed in 20 ul volumes of 1.5 mM $MgCl_2$, 5 mM NH_4Cl, 10 mM Tris (pH8.3), 50 mM KCl, 100 μM dNTPs, with 1 U Taq DNA polymerase and 1μM of each primer. Primers used were sY581f- CACTGCCCTAATCCTAGCACA, sY581r-TCTTCTGGACATCCACGTCA; sY586f-GTGTGGCACATATGCCTATAAA, sY586r-TTGGTACATCCAGATGCAGAT; sY587f-TGGTTAATAAAGGGAAGGTGTTTT, sY587r-TCTCCAGGACAGGAAAATCC. Thermocycling conditions consisted of an initial denaturation of 3 minutes at 94°C; 35 cycles of 1 minute at 94°C, 1 minute at 64°C,1 minute at 72°C; and, finally, 5 minutes at 72°C. Amplification products were digested with Sau3A (sY581), TaqI (sY 586) or DraI (sY587) for 3 hours and run on 4% Nusieve agarose gels.

Results and Discussion

Breakpoint of an *AZFc*-deleted man -WHT 2564

WHT 2564 is an infertile man with azoospermia. Testis biopsy revealed an equal distribution of tubules with only sertoli cells (no germ cells), and with maturation arrest (spermatocytes arrested in meiosis I). No sperm were detected in testis biopsies. Y chromosomal STS analysis on DNA from WHT 2564 reveal an *AZFc* region deletion (Fig 1). Some STSs that are commonly deleted in *AZFc* deleted men are present on the Y chromosome of WHT 2564, suggesting that WHT2564 has a smaller deletion than other men without the *AZFc* region.

We mapped the proximal deletion breakpoint of WHT 2564 by analyzing all sequence family variants and microsatellite repeats known to map within and around the *AZFc* region by PCR-SSCP or PCR-sequencing. The results of this analysis are shown in Figure 2. WHT 2564 had only one sequence variant associated with some *AZFc* region SFV-STSs (including sY790) and tested negative for all *AZFc* region STS's distal to sY 790 (including sY813). STSs sY 790 and sY 813 mapped within the same sequenced Y BAC clone RP11-289L7, and were 24 kb apart from each other, and only 37 kb upstream of the most proximal DAZ gene (*DAZ1*). Within the intervening sequence lies a 12 kb HERV element (human endogenous retroviral element) that is a likely candidate for the recombination event mediating this deletion. A 10kb retroviral element, with little homology to this retroviral element, is responsible for the *AZF*a deletion seen in two patients (C. Sun, in preparation). We designed STSs proximal to, distal to and within this element and tested DNA from WHT 2564 for the presence or absence of these STS's. We found that sequences immediately distal to this element were absent in WHT 2564. Confirmation of the distal breakpoint awaits full sequence of an *AZFc* distal region BAC in which a similar retroviral element lies. We have however limited the distal breakpoint for this patient to within the euchromatin on Yq, in a region that is highly homologous to subtelomeric repeats in other chromosomes. This deletion spans approximately 3.5 Mb. Within the deleted region are the four known *AZFc* region gene families: two copies of BPY2, four copies of DAZ, two copies of *CDY1* and several RBM genes. This deletion breakpoint is rare, as WHT 2564 is the only man

we identified who has a deletion retaining a copy of BPY2 but in whom all copies of DAZ and *CDY1* are deleted. No other such patients have been reported in the literature.

WHT 2564 has the most severe *AZFc* phenotype, although his deletion is the smallest observed within *AZFc* so far. This forces us to postulate factors outside *AZFc* as exacerbating the spermatogenic failure phenotype in this individual as compared to other *AZFc* nulls. There is no record of the age at which the testis biopsy on WHT 2564 was performed, and it is possible that this man, like other *AZFc* nulls, had an infertility phenotype that progresses with age.

The *de novo AZFb* region deletion in WHT 3097 extends into *AZFc*

WHT 3097 is deleted for *AZFb* region STSs corresponding to specific copies of RBM (L. Brown, unpubl., Figure 1). Analysis with DAZ STSs that detect sequence family variants in different DAZ genes (sY581,s586, sY587) revealed that WHT 3097 also had a de novo deletion of DAZ genes *DAZ1* and *DAZ2* in addition to an RBM region deletion (not shown). PCR-SSCP and PCR-sequencing as above revealed that 3097 has a breakpoint in a region of repeated sequence between the DAZ gene clusters (figure 2 WHT 3097, WHT 3244 father).

Although the sequences at the breakpoint mediating this deletion have not yet been identified, this result raises an interesting question. Are all *AZFb* region deletions actually deletions that overlap with the *AZFc* region, and lower the copy number of multiple *AZFc* genes as well? These questions could not be answered by conventional STS analysis, as no *AZFc* region markers are actually deleted in WHT 3097, only the copy number is reduced. Assays that detect variant nucleotides within a sequence family have permitted unmasking of a potential *AZFc* region contribution to the phenotype of *AZFb* deleted men. We have no other patients in our collections with *AZFb* region deletions, but our collaborators in Holland (Hoovers, JM and colleagues) independently identified such a patient by FISH using DAZ gene probes. DNA analysis with WHT 3947, this second *AZFb*/c deleted patient, confirms that this man has a distal deletion breakpoint within the same large block of repeats as WHT 3097. Analysis of *AZFb* region markers revealed that the proximal breakpoint in this patient is more proximal to the Y centromere than in WHT 3097.

The genes deleted in WHT 3097 are multiple copies of *RBM* (and all copies within *AZFb* which are the most functionally active)(Elliott et al. 1997), one copy of *BPY2* and two copies of

DAZ. WHT 3097 is severely oligospermic, with a sperm count of < 1million/ cc. Some semen samples from this individual had no sperm at all. *AZFb* deleted men are usually azoospermic and exhibit maturation arrest, but the *AZFb* deletion in this man is slightly smaller than other reported *AZFb* deletions. X-Y homologous genes *EIF1AY* and *SMCY* lie in the part of the *AZFb* region that is not deleted in WHT 3097, so the possibility that they are responsible for the more severe phenotype in *AZFb* deleted men should be explored.

CDY1 deletions encompass one *DAZ* cluster

WHT 3509

Infertile male WHT 3509 and his father, WHT 3518, have deletions of the *CDY1* gene (B. Lahn, unpublished results; Figure 1). As the *CDY1* genes flank the DAZ cluster consisting of genes *DAZ3* and *DAZ4*, we tried to map the breakpoint of these men as well. Our results are shown in figure 2 and Table 2. The proximal breakpoint for these men lies distal to sY783. The distal breakpoint lies within the euchromatin. The deletion spans at least 2Mb. WHT 3509 is infertile, with azoospermia. No more information is available for this patient. Known genes deleted in this pair of men are two copies of *CDY1* and *DAZ3* and *DAZ4* . Although WHT 3518 is fertile, it is possible that he is sub-fertile, and fathered a son before becoming azoospermic, as reported for other *AZFc* deleted men (Chang et al. 1999). Sperm counts of WHT 3518 could still be tested to confirm this possibility. Examination of the Y haplotype of these men revealed that they are in haplogroup X. Other men from this haplotypic group could be tested for *CDY1* deletions to examine the stability of this chromosome in the population.

These results show that deletions of the *AZFb* and *AZFc* regions in infertile males are heterogenous, and can even be overlapping. The distinction between traditionally defined *AZFb* and *AZFc*, which were each considered to encompass only certain candidate genes, may not actually exist. Instead, as seen with the three patients described here, each individual's deletion may be an independent combination of reduction in copy number of some genes, and absence of other genes.

Analysis of breakpoints in these and other cases will point to the mechanisms of *AZFc* region deletions. In both *AZF* deletion breakpoints characterized thus far, in this report (WHT2564) and in the *AZF*a region breakpoint studied by Sun et al (pers. comm.), sequences

mediating recombination have been within human endogenous retroviral elements. This comes from a large family of sequences, which comprises about1% of the genome(Sverdlov 2000). Features of these elements, and even their possible transcription in male germ cells during spermatogenesis may make intrachromosomal or sister chromatid recombination between these elements, and resulting deletion of intervening sequences possible.

Correlation of genotypes with phenotypes in *AZF* deleted men has been very difficult, and analysis of this nature, which involves characterizing deletion breakpoints in many infertile men will be crucial in understanding how AZF region deletions cause male infertility.

References

Chang PL, Sauer MV, Brown S (1999) Y chromosome microdeletion in a father and his four infertile sons. *Human Reproduction* **14**:2689-94.

Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, et al (1997) Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proceedings of the National Academy of Sciences* **94**:3848-3853.

Mulhall JP, Reijo R, Alagappan R, Brown L, Page D, Carson R, Oates RD (1997) Azoospermic men with deletion of the *DAZ* gene cluster are capable of completing spermatogenesis: Fertilization, normal embryonic development and pregnancy occur when retrieved testicular spermatozoa are used for intracytoplasmic sperm injection. *Human Reproduction* **12**:503-508.

Reijo R, Alagappan RK, Patrizio P, Page DC (1996) Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome. *Lancet* **347**:1290-1293.

Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, et al (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genetics* **10**:383-393.

Silber SJ, Alagappan R, Brown LG, Page DC (1998) Y chromosome deletions in azoospermic and severely oligozoospermic men undergoing intracytoplasmic sperm injection after testicular sperm extraction. *Human Reproduction* **13**:3332-3337.

Sverdlov ED (2000) Retroviruses and primate evolution. *Bioessays* **22**:161-171.

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, et al (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics* **5**:933-943.

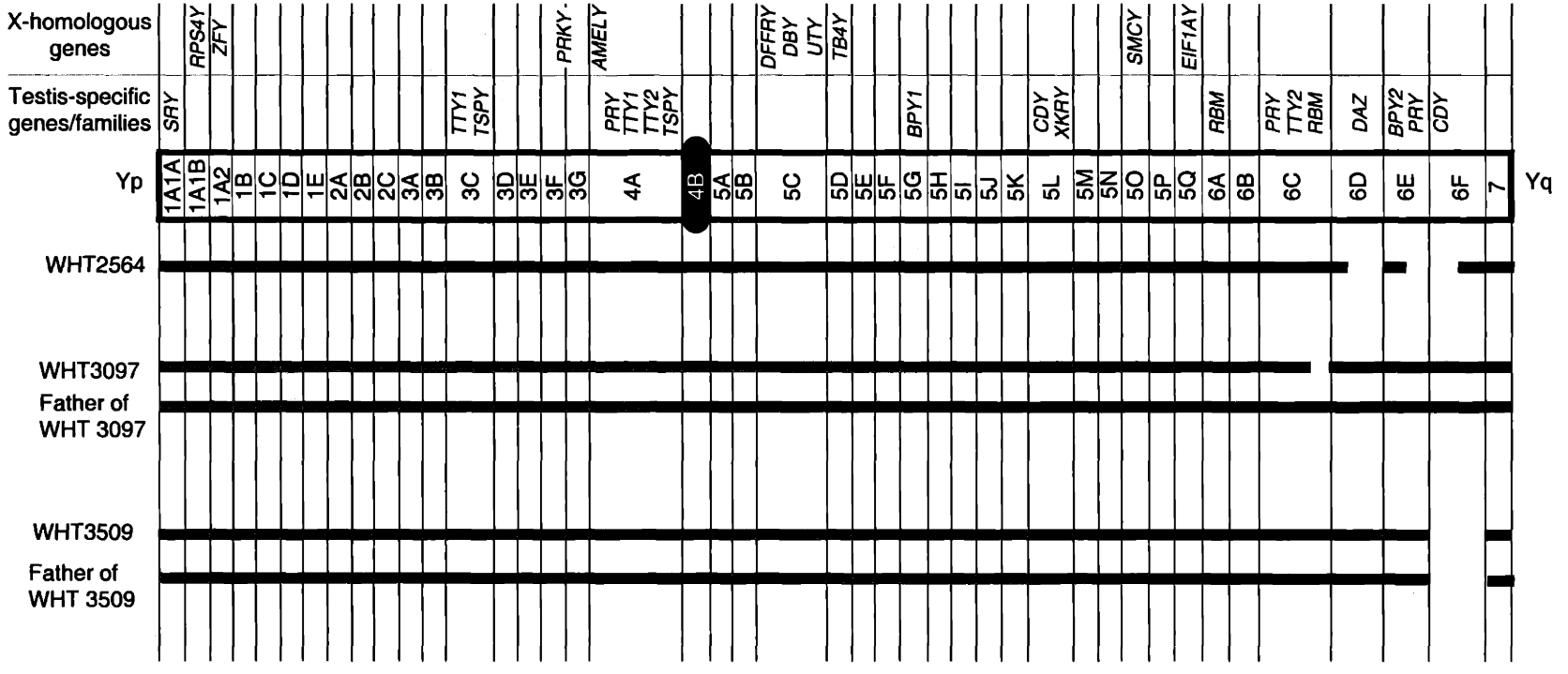


Figure 1. STS deletion profiles of three infertile male patients, WHT 2564, WHT 3097, and WHT 3509. STS based map of the Y chromosome shows Y deletion intervals 1 through 7, and all known NRY gene families. Solid bar shows positive STS markers, whereas gap shows regions where markers tested negative.

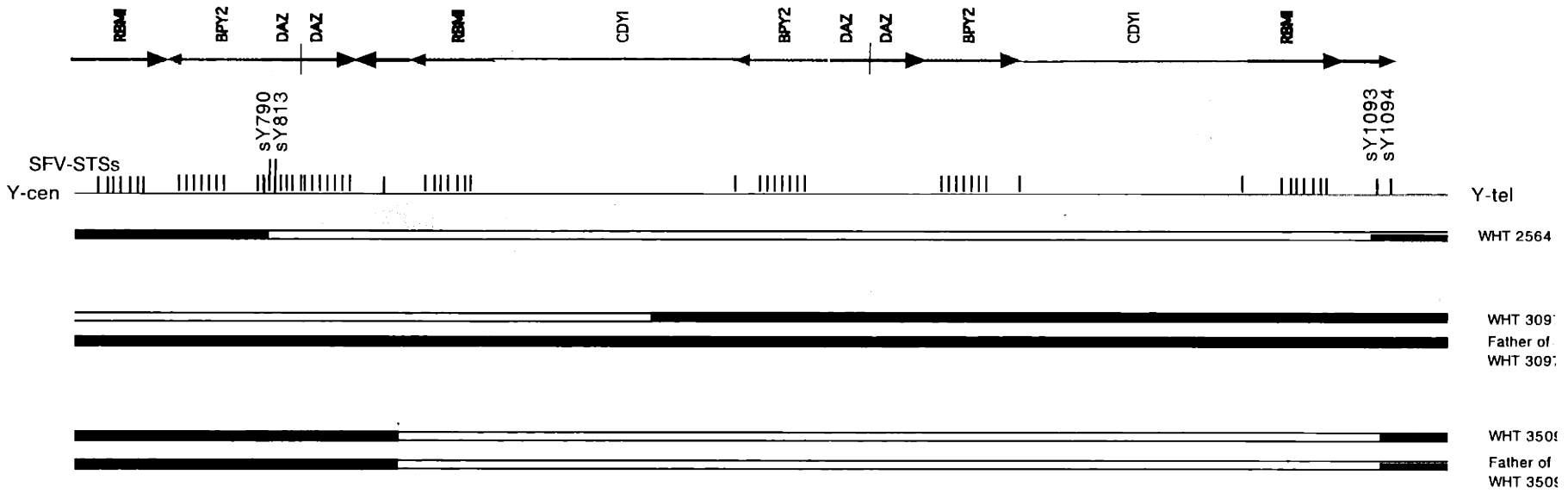


Figure 2. Fine genomic organization of the *AZFc* region. Colored bars represent large blocks of repeated sequences. Genes within each repeat are indicated. (courtesy Y chromosome sequencing project, H. Skaletsky, T. Kawaguchi and D.C. Page). Key STS's that delimit the breakpoints in each male are labeled, other STSs tested are indicated by vertical lines. Filled bars represent regions testing positive for each male, and empty regions indicate

200kb

Appendix 2

Mammalian DAZ proteins show rescue of the meiotic entry defect of *Drosophila boule*

R.Saxena and J. Maines

Introduction

Four *DAZ* genes lie within the *AZFc* region in most human Y chromosomes, and they are likely contributors to the phenotype of *AZFc* deleted men. Transcripts have been observed from at least three genes, and these or even all four genes may be translated (Chapter 3). The *DAZ* genes encode putative RNA binding proteins, and are predominantly expressed in spermatogonial stem cells of the testes, and in spermatocytes(Menke et al. 1997; Niederberger et al. 1997). DAZ proteins are most abundant in spermatocytes, the cell types that undergo meiosis (R. Reijo et al, pers. comm.). The *DAZ* gene family in humans also consists of an autosomal gene *DAZL*, which is closer in identity to an ancestor from which the *DAZ* genes arose(Saxena et al. 1996). DAZ proteins have tandem amplification of one (*DAZ2,4*) or two (*DAZ1,3*) internal regions that result in multiple RRM domains in proteins from *DAZ1* and *DAZ3* and multiple copies of a 24 aa unit of unknown function in the C terminal ends of all four proteins. DAZL protein does not have any internal repeats.

DAZ gene family homologues have been described in many different organisms after the identification of *DAZ* genes in humans in our lab in 1995 (Eberhart et al. 1996, Reijo et al. 1996, Houston et al. 1998, Gromoll et al. 1999, Karashima et al. 2000;). Y linked *DAZ* homologues only exist in old world monkeys, and no mutations in *DAZ* alone have been identified in humans or in other species with Y-linked *DAZ* genes. Therefore, there is no direct evidence of a role of DAZ in spermatogenesis. The presence of multiple, similar copies of *DAZ* in humans has made hunts for point mutations or deletions limited to *DAZ* in infertile men difficult. Furthermore, if all four genes are functional, the disruption of only one gene may result in a phenotype too mild to detect in patient populations. As no mutations have been found in *DAZ* alone, and the *DAZ* genes appear to be evolving rapidly, the candidacy of *DAZ* as an azoospermia factor has been questioned by Agulnik et al, who have claimed that *DAZ* may have little or no role in spermatogenesis(Agulnik et al. 1998).

On the other hand, genetic analysis of *DAZL* homologues has been ongoing in both vertebrate and invertebrate model systems and suggests that this gene family has critical germ cell functions in multiple organisms. Mouse *dazl* is required for male and female germ cell survival during embryogenesis (Ruggiu et al. 1997). *Xenopus xdazl* is required for early primordial germ cell differentiation(Houston and King 2000). *boule*, the *Drosophila DAZ*

homologue, is required for meiosis in males(Eberhart et al. 1996). *C. elegans-dazl1* is required for meiosis in the hermaphrodite germ line(Karashima et al. 2000).

The roles of human DAZ proteins in different aspects of germ cell development could be studied by complementation of mutant phenotypes using ectopically expressed human DAZ proteins. This is a feasible strategy as the DAZ proteins are well conserved at the amino acid level during evolution, and have performed germ cell functions at least since the time of divergence between the human and nematode lineages. Experimental evidence already exists for some conservation of DAZ function in different organisms. In 1998, Houston et al found that *Xenopus xdazl*, when ectopically expressed in spermatocytes of flies, could rescue the meiotic entry defect of *Drosophila boule* to the same extent as ectopic expression of boule using the same expression system(Houston and King 2000). Recently, Slee et al described that introduction of a YAC clone containing exons 2-11 of one DAZ gene and about 150kb of surrounding sequence into *dazl-/-* mice resulted in partial rescue of the mutant phenotype, manifest as a pronounced increase in spermatogonial germ cells in the adult, and some survival to pachytene stages of meiosis. The results from this experiment should be interpreted with caution because the 5' end of the *DAZ* gene, including the promoter region, and exon 1 (5' untranslated region and the start codon of DAZ) were not included in the transgene, although the authors report some testis specific expression of a chimaeric DAZ protein (Slee et al, 1999).

In order to test whether mammalian DAZ proteins could functionally compensate for *boule* in meiosis, and specially to test whether any of the highly amplified Y-encoded proteins were capable of this germ cell function, we expressed the human DAZ and mouse DAZL proteins in flies using a testis specific expression system. We crossed these flies into a *boule* mutant background, and assayed the extent of meiotic rescue using immunohistological criteria.

Materials and Methods

Cloning of expression constructs

To generate the human *DAZ* rescue construct, a1.6 kb DraIII/BamHI fragment of human *DAZ* cDNA pDP1678 containing the entire coding sequence of *DAZ* was ligated by blunt-end ligation into the EcoRI site of the β2-tubulin testis vector(Hoyle et al. 1995). The β2-tubulin testis vector is a P-element transformation vector carrying a w+minigene, and promoter and

regulatory sequences from the β2-tubulin gene. Clones in the correct orientation were selected by PCR using a forward primer from the β2-5'UTR and a reverse primer from within DAZ coding sequence. PCR products from both junctions were sequenced to ensure integrity of the DAZ orf. To generate the mouse *Dazl (mDazl)* construct, primers with EcoRI linkers were used to PCR amplify the coding sequence from mouse *Dazl* cDNA clone pDP1580, and EcoRI digested PCR products were ligated into the EcoRI site of the β2-tubulin testis vector. Cloned inserts in the correct orientation were completely sequenced.

Germline Transformation

Rescue constructs were co-injected with a helper plasmid phsD2-3, ry+ (containing transposase) into w^{1118} embryos using standard procedures. (Spradling 1986). Two transgenic lines, one on chromosome 2, and one on the X chromosome were generated for mDazl, and eleven transgenic lines were generated for hDAZ2, of which four were randomly chosen for characterization. *Boule* is located on chromosome 3. Genetic crosses were carried out to generate *boule* homozygotes carrying one copy of either the mDAZL transgene or one copy of the hDAZ transgene.

Analysis of fixed Drosophila testes contents

Testes from newly eclosed, 0 to 1 day old males were dissected in Ringer's solution, fixed and flattened under a coverslip for analysis by phase microscopy. Fixation and indirect immunoflourescence was done according to the method of Cenci et al (1994). The M-phase marker phosphorylated histone H3 was detected with a polyclonal antiserum (Upstate Biotechnologies, 1:200). Fluorescently labeled secondary antibodies (Jackson Labs) were used at 1:500. DNA was also stained with 1mg/ml DAPI. Fluorescent images were recorded from a Leica DMRXE microscope equipped with a charge-coupled device (CCD) digital imaging camera (DAIC; Hamamatshu)

Results and Discussion

Expression of mammalian DAZ proteins in Drosophila testes

The coding sequences of mouse *Dazl* (mDazl), and human *DAZ* (hDAZ2; from gene *DAZ2*) were expressed under the control of the spermatocyte-specific β2 tubulin promoter and the 5' and 3' untranslated sequences of the β2-tubulin gene. The hDAZ2 protein has one RNA binding domain and 15 tandemly repeated 24 amino acid units (Chapter 3, figure 6). The β2 tubulin gene encodes a testis-specific isoform of b-tubulin that is required for all microtubule based functions in meiotic and postmeiotic germ cells. β2 expression is first detected in late primary spermatocytes, and this expression system mirrors that expression pattern for all proteins placed within it. (p109, Fuller 1993)(Hoyle et al. 1995). Multiple transgenic lines were generated by P-element mediated germ-line transformation. Integrity of the mDazl and hDAZ2 orf within transgenic flies was confirmed by PCR on fly genomic DNA. Expression of the transgene was detected by RT-PCR in mDazl transgenic males. Fertility of all transgenic lines did not appear to be impaired.

Rescue of meiotic entry defect of boule by mouse and human DAZ proteins

boule is required for entry of spermatocytes into meiosis. Expression of boule cDNA in mutant flies using this expression system results in rescue of meiotic entry in over half the germline cysts. Fertility is however, not restored in these flies (Eberhart et al. 1996). Crossing mDazl and hDAZ transgenic flies into the boule background too resulted in a partial rescue of the boule meiotic entry defect (Figures 1, 2, 3).

Meiotic entry was assayed by appearance of phosphohistone H3-positive chromosomes, and movement of these condensed chromosomes to the metaphase plate. This M phase marker shows most clearly the stages of chromosome condensation and movement that characterize the transition between prophase to metaphase of meioisis I. Figures 1a, b and overlap of images in c are from meiotic cysts from boule deficient flies carrying the *mDazl* transgene. The cyst to the left of the picture is entering the metaphase transition of meiosis. Phosphohistone H3 (PH3) staining overlaps with three visibly condensed chromosomes in most cells within the 16 cell cyst. Within this cyst, all spermatocytes are almost in synchrony, but the trio of chromosomes in the

middle of the picture (staining with DAPI but not with PH3, figure 1c) are not as close to each other as the others, and are lagging behind. Meiotic entry occurs as a wave and is not completely synchronous within a cyst; that cell has not passed the transition to metaphase. In this squash, on the right are some primary spermatocytes that have not yet entered meiosis. Chromosomes are condensed and stain strongly with DAPI but do not stain with phosphohistone H3 (figure 1a, b, c).

The controls, in figures 1d, e and f, show late primary spermatocytes in boule homozygotes. In boule homozygotes, the cell cycle in spermatocytes arrests in the transition to metaphase of meiosis I. No phosphohistone H3 staining is coincident with chromosomes, although there is some background PH3 staining.

Figures 2a and 2b show PH3-positive chromosomes in more cysts from mDazl transgenic, boule deficient flies. Chromosomes are in different stages of migration to the metaphase plate.

Figure 3 shows PH3 staining of a cyst from boule homozygotes carrying the hDAZ2 transgene. Here too, chromosomes stain brightly, and in most cells of the cyst, the chromosomes have moved to the metaphase plate for meiosis I. Our results suggest that the hDAZ2 protein, despite having over 300 additional amino acids, and fourteen DAZ repeats more than mDAZL (298 aa) or boule (228 aa) is still capable of rescuing the meiotic defect of boule deficient males. Cells in anaphase of meiosis I and undergoing cytokinesis will be more evident by staining of the spindle apparatus with anti-tubulin antibodies, however this has not yet been done.

Preliminary evidence suggests that the hDAZ2 transgenic lines (all four tested) have more cysts entering meiosis than mDAZL transgenic lines (two tested), which is surprising. Further experiments will be needed to determine why there was a greater extent of rescue with hDAZ2 than mDAZL. Similar experiments using hDAZL transgenes are underway, and it would be interesting to see how well these transgenes would work under this expression system. This experiment should be repeated in a more rigorous way. Ectopic expression of human DAZ and DAZL proteins could be driven by the endogenous boule promoter to recapitulate boule expression, and genomic sequence of boule is now available to design such a construct. All four human Y linked DAZ genes could be expressed in flies and subtle differences between efficacy of the proteins to function in meiosis examined. Expression of other RNA-binding proteins using

this same expression system has not led to rescue of the boule phenotype. Pelota, a translation factor, was ectopically expressed in a similar manner, and it showed no ability to restore meiotic entry in boule mutants. (J. Maines, pers. comm.).

This is the first convincing demonstration that DAZ genes are capable of function, and suggests that the absence of DAZ in the AZFc deleted men may affect meiotic entry in spermatocytes. This role of DAZ family proteins in meiosis appears to be the most conserved. Both *C. elegans* and *Drosophila* mutants in *DAZ* homologous genes have sex-specific defects in meiotic entry even though the sexual development process affected is opposite in each case(Eberhart et al. 1996; Karashima et al. 2000), spermatogenesis in male flies and oogenesis in hermaphrodite worms. Although the mutant phenotypes first seen in frogs lacking *xdazl* and mice lacking *dazl* are in spermatogenic precursors which would have to undergo multiple rounds of mitosis and differentiation before meiosis, it is of interest that the highest level of protein expression in both species is in spermatocytes, just as they are about to enter meiosis. It is possible that the role of DAZ proteins in meiosis in these species is masked by additional roles in earlier germ cell development. Houston et al's demonstration of the capacity of Xdazl to rescue boule meiotic defects would suggest that this is true (Houston et al. 1998). Recent studies in the rat confirm that Dazl protein expression in that species is highest in the cytoplasm of pachytene spermatocytes (Rocchietti-March et al. 2000). Human DAZ and DAZL proteins are also most abundant in primary spermatocytes (R. Reijo et al, submitted). The experiment presented here opens the way for analysis of the role of human DAZ proteins in meiosis - one crucial step during germ cell differentiation.

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. *Human Molecular Genetics* 7:1371-1377.

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

Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, et al (1997) Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proceedings of the National Academy of Sciences* 94:3848-3853.

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. *Human Molecular Genetics* 8:2017-24.

Houston DW, King ML (2000) A critical role for Xdazl, a germ plasm-localized RNA, in the differentiation of primordial germ cells in Xenopus. *Development* 127:447-456.

Houston DW, Zhang J, Maines JZ, Wasserman SA, King ML (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.

Hoyle HD, Hutchens JA, Turner FR, Raff EC (1995) Regulation of beta-tubulin function and expression in Drosophila spermatogenesis. *Developmental Genetics* 16:148-170.

Karashima T, Sugimoto A, Yamamoto M (2000) Caenorhabditis elegans homologue of the human azoospermia factor DAZ is required for oogenesis but not for spermatogenesis. *Development* 127:1069-1079.

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

Menke DB, Mutter GL, Page DC (1997) Expression of *DAZ*, an azoospermia factor candidate, in human spermatogonia. *American Journal of Human Genetics* 60:237-241.

Niederberger C, Agulnik AI, Cho Y, Lamb D, Bishop CE (1997) In situ hybridization shows that *Dazla* expression in mouse testis is restricted to premeiotic stages IV-VI of spermatogenesis. *Mammalian Genome* 8:277-278.

Reijo R, J. S, 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.

Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, et al (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genetics* 10:383-393.

Rocchietti-March M, Weinbauer GF, Page DC, Nieschlag E, Gromoll J (2000) Dazl protein expression in adult rat testis is up-regulated at meiosis and not hormonally regulated. *International Journal of Andrology* 23:51-56.

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

Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, et al (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.

Slee R, Grimes B, Speed RM, Taggart M, Maguire SM, Ross A, McGill NI, Saunders PT, Cooke HJ. (1999) A human DAZ transgene confers partial rescue of the mouse Dazl null phenotype. *Proc Natl Acad Sci U S A.* Jul 6;96(14):8040-5.

Spradling A (1986) P element-mediated transformation. In: Roberts DB (ed) Drosophila: A Practical Approach. IRL Press, Oxford, pp 175-197

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, et al (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics* 5:933-943.

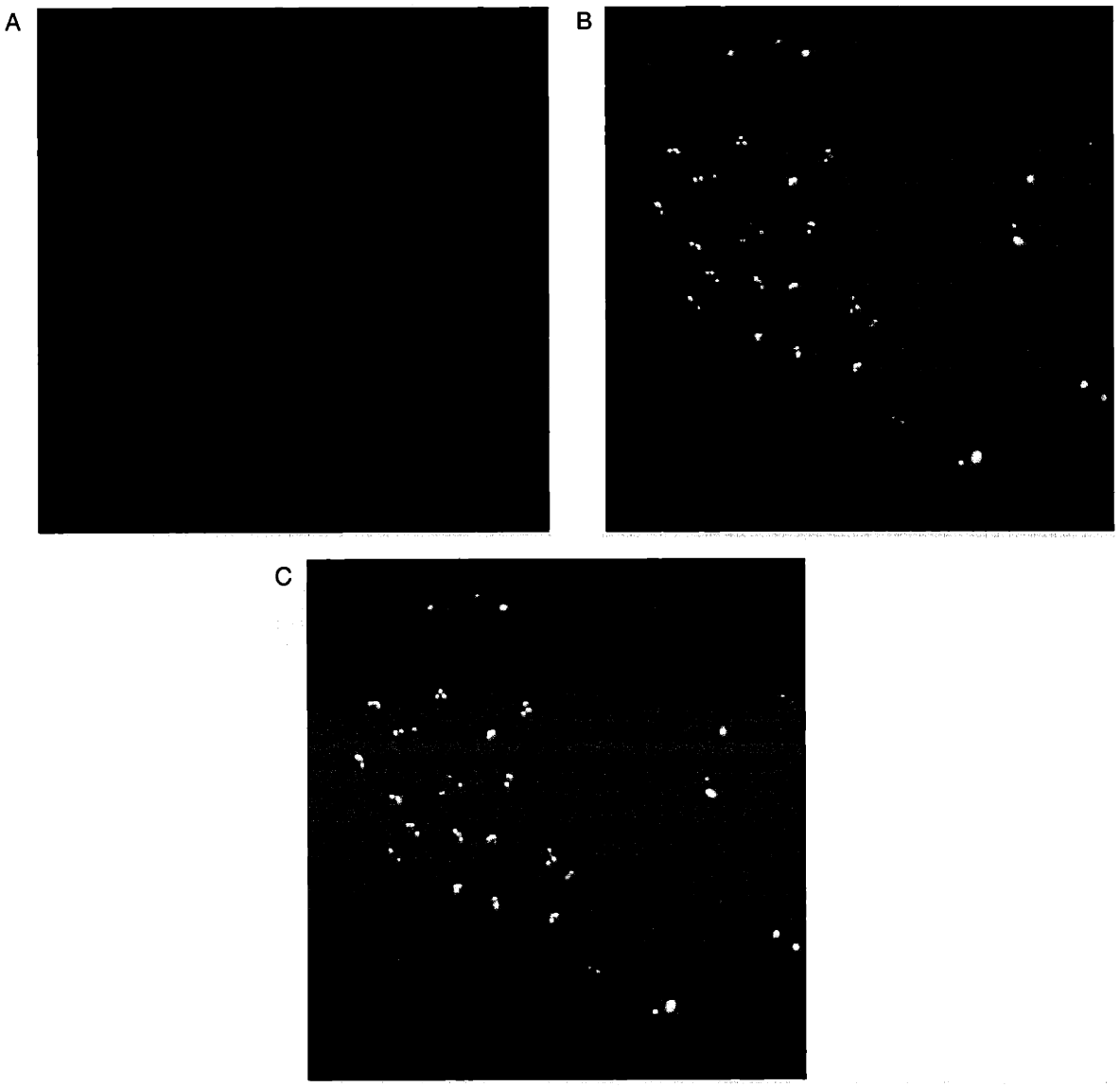


Figure 1. *a b c* Expression of mDAZL rescues meiotic entry in *boule* mutant testes. Photographs of squashed, fixed testis from 0-1day old boule homozygous males carrying the mDAZL transgene A) Phosphohistone H3 staining of spermatogenic cysts from mDAZL; bol-/ bol- males, B) DNA visualized by DAPI staining of same spermatogenic cysts. C. Overlay of phosphohistone H3 and DAPI images

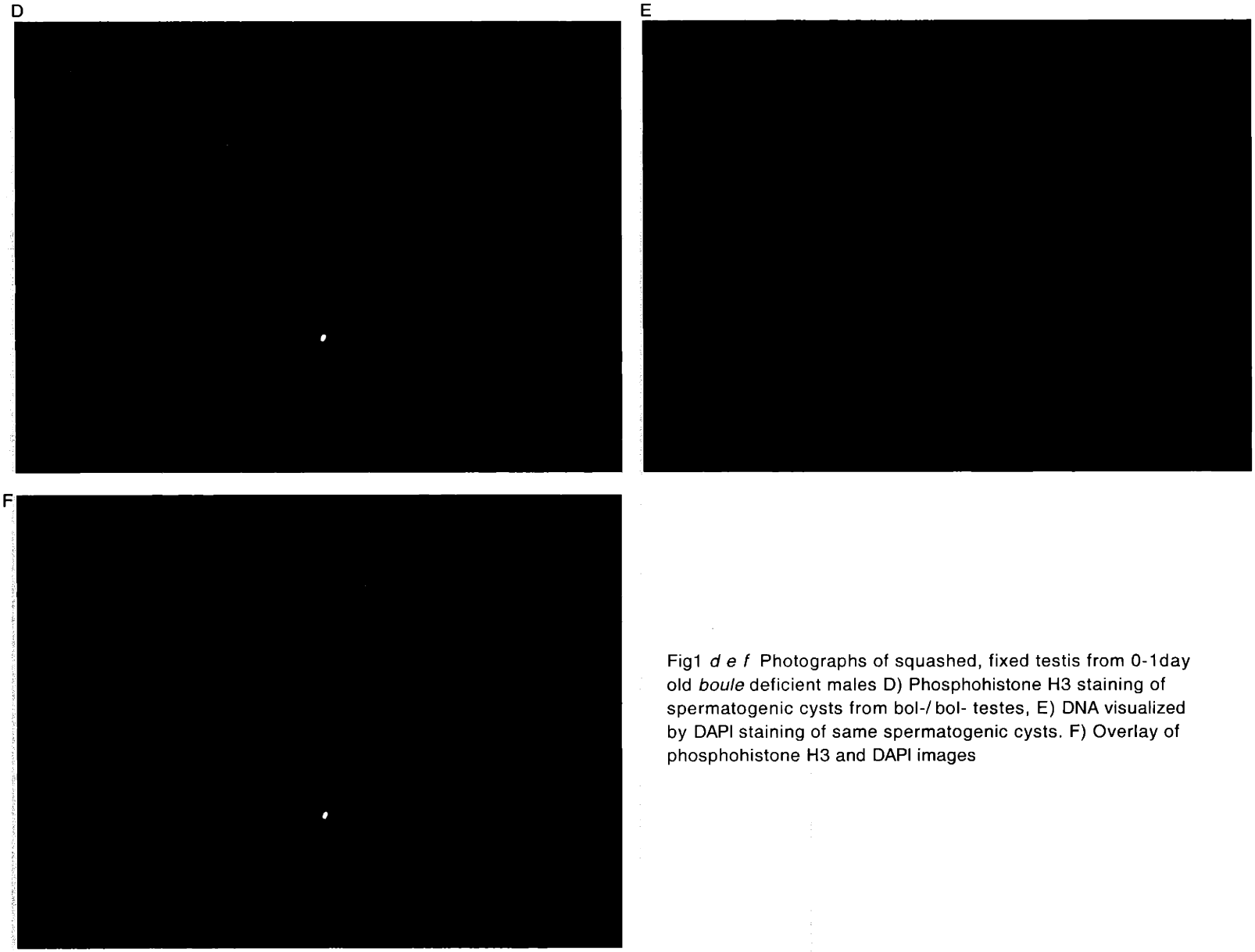
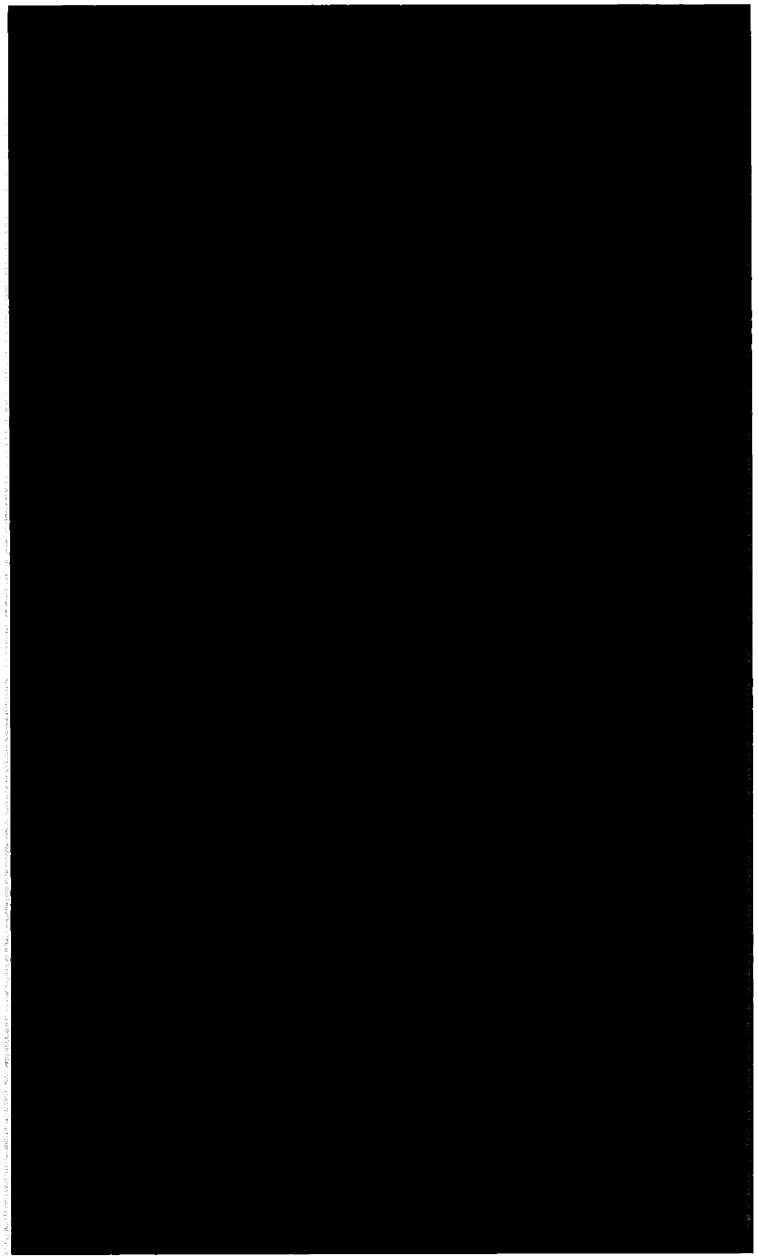


Fig1 *d e f* Photographs of squashed, fixed testis from 0-1day old *boule* deficient males D) Phosphohistone H3 staining of spermatogenic cysts from bol-/ bol- testes, E) DNA visualized by DAPI staining of same spermatogenic cysts. F) Overlay of phosphohistone H3 and DAPI images



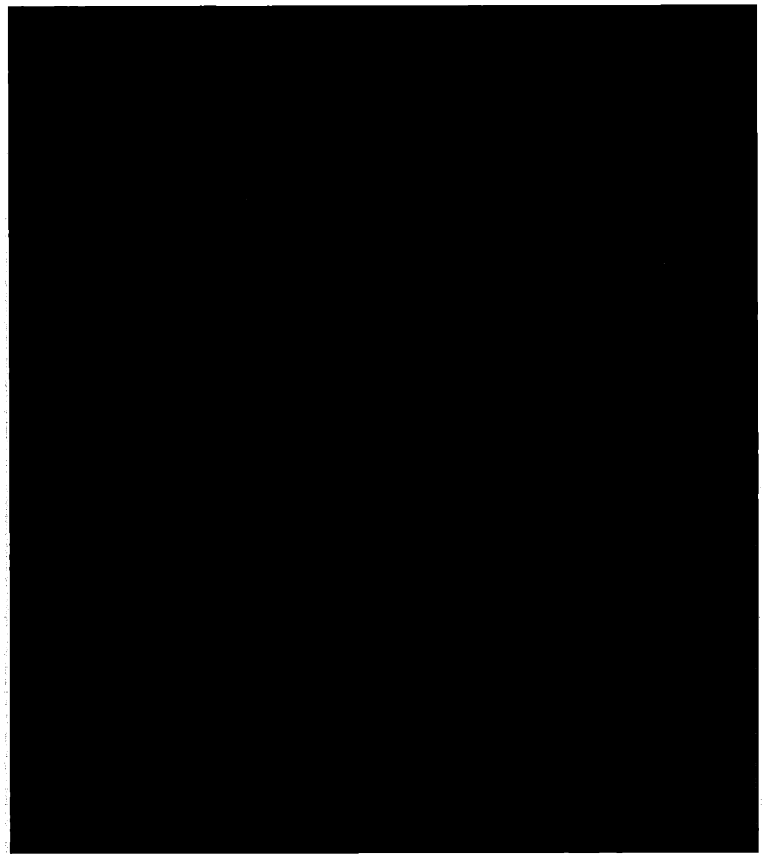


Figure 2. *a b* Rescue of *boule* meiotic entry defect by mDAZL. a,b Phosphohistone H3 staining of chromosomes in cysts from *boule* deficient flies carrying the *mDazl* transgene. Chromosomes are in the process of coming together at the metaphase plate.

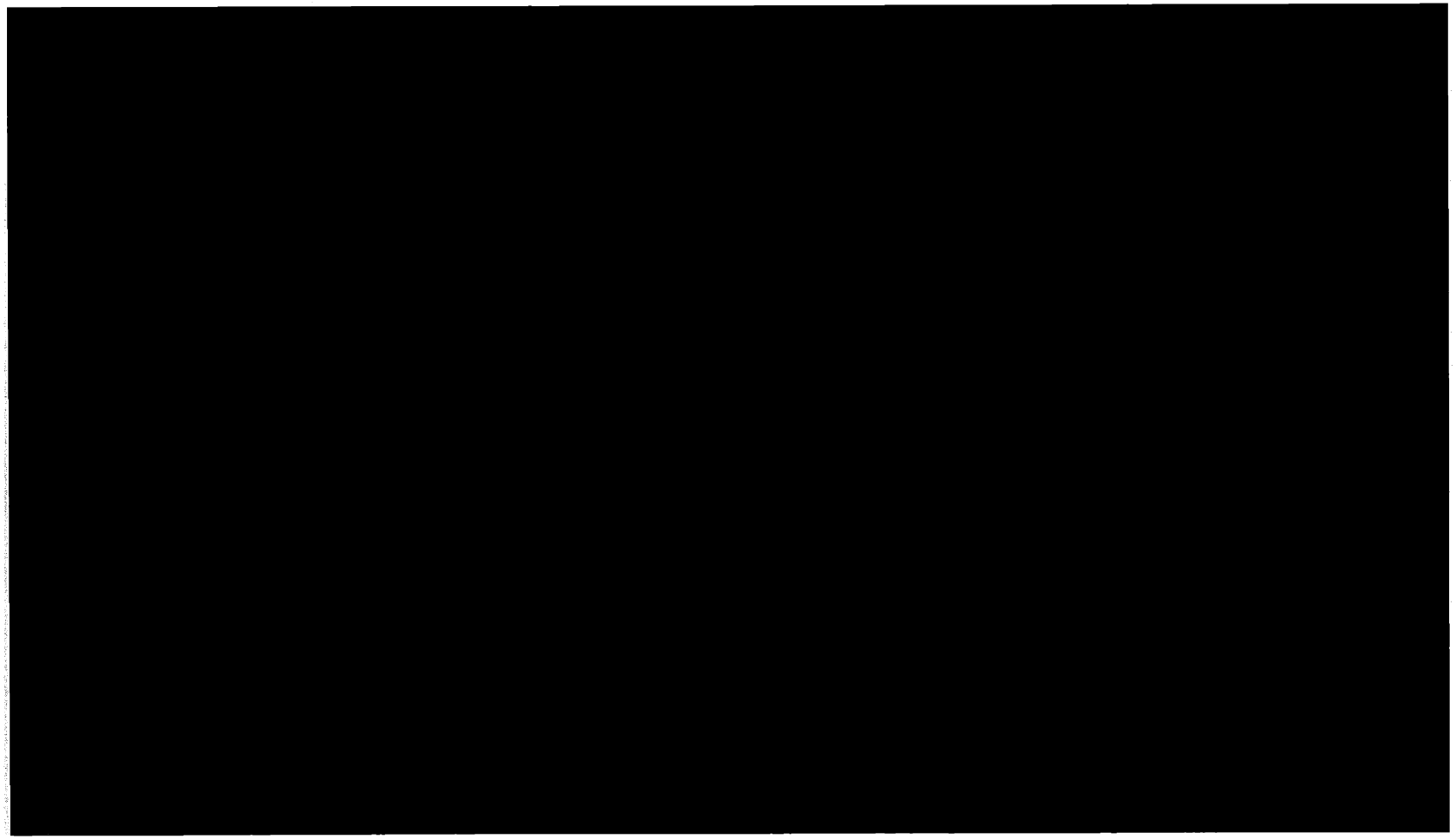


Figure 3. Rescue of *boule* meiotic entry defect by hDAZ2. Phosphohistone H3 staining of chromosomes in spermatogenic cyst from *boule* deficient fly carrying the *hDAZ2* transgene. Chromosomes are moving towards, or are at the metaphase plate.