Genomics **67,** 256–267 (2000)
doi:10.1006/geno.2000.6260, available online at http://www.idealibrary.com on IDEAL®

Four *DAZ* Genes in Two Clusters Found in the *AZFc* Region of the 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[1]

*Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142; *Center for Reproductive Medicine, Department of Gynecology and Obstetrics, and ‡Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands; and †Celera Genomics, Foster City, California 94404*

Received March 22, 2000; accepted May 25, 2000

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.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. G63906 (sY581), G63907 (sY586), G63908 (sY587), G63909 (sY579), G63910 (sY776), AF271087 (pDP1575), AF271088 (pDP1576) AF248480 (pDP1678), AF248481 (pDP1679), AF248482 (pDP1680), and AF248483 (pDP1681).

[1] To whom correspondence should be addressed. Telephone: (617) 258-5203. Fax: (617) 258-5578. E-mail:dcpage@wi.mit.edu.

INTRODUCTION

Approximately 2% 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 *et al.* (1997) found evidence of multiple *DAZ* genes on Yq by fluorescence *in situ* hybridization (FISH). By Southern blotting and long-range restriction mapping, Yen and

0888-7543/00 $35.00

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 *et al.* (1998) observed what they interpreted to be seven *DAZ* genes or pseudogenes on Yq.

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 10 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 RP11-290O3 (GenBank Accession No. AC010089). As shown in Fig. 1, cosmid 18E8 encompasses the 5′ portions of two neighboring *DAZ* genes. Cosmid 63C9 (Saxena *et al.*, 1996; GenBank Accession No. AC000021) contains exons 2 through 11 and thus almost an entire *DAZ* gene. Cosmid 46A6 (Saxena *et al.*, 1996; GenBank Accession No. AC000022) derives from the 3′ portion of *DAZ;* it contains exons 8 through 11 as well as 35 kb downstream of the gene.

Fluorescence in situ hybridization. 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 being blow-dried, slides were incubated in extraction solution (0.125% SDS, 0.2 M 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 the 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 preheated 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, 165 kb; 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, 200 kb; 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 (pDP1649; 2.4-kb insert from the *DAZ* genomic locus). Probes were labeled with [^{32}P]dCTP by random priming. Hybridization was carried out at 65°C in 0.5 M $NaPO_4$ (pH 7.5), 7% SDS. Membranes were subsequently washed at 65°C in 0.1× 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 (pH 8.3), 50 mM KCl, 100 μM dNTPs, with 1 U *Taq* DNA polymerase and a 1 μM concentration of each primer. PCR primers and conditions have been deposited with GenBank under the following accession numbers: sY581, G63906; sY586, G63907; sY587, G63908; sY579, G63909; and 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). λ 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 was 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, 5 mg/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 of 7.5 M ammonium acetate, 75 μl isopropanol) and washing with 70% ethanol.

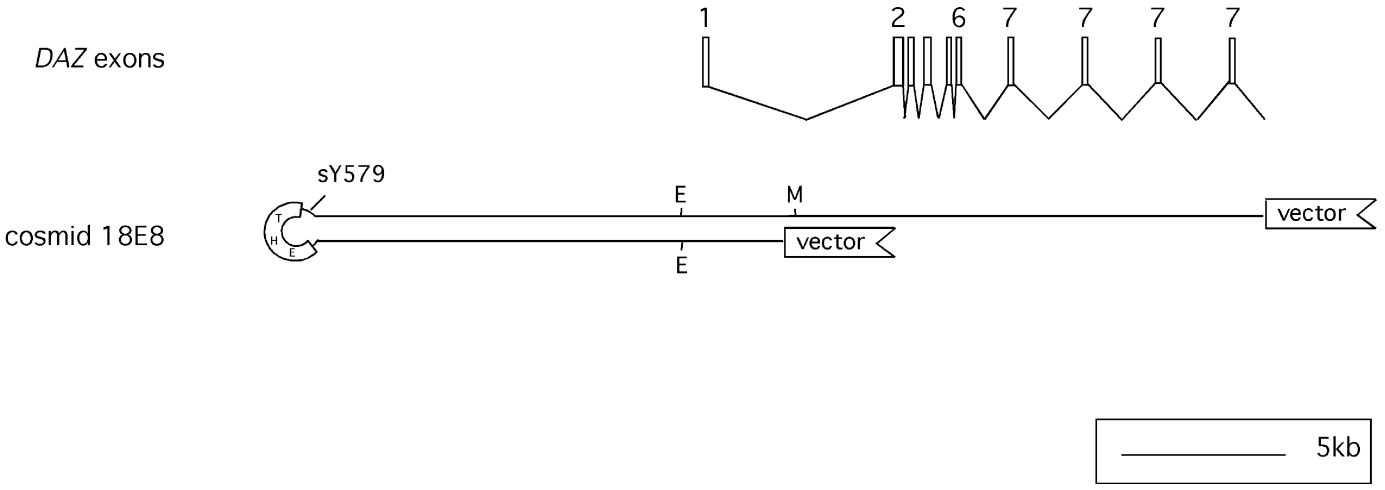

FIG. 1. Schematic diagram of inverted duplication in cosmid 18E8. *DAZ* exons are shown at the top. A small nonduplicated segment containing a 1.9-kb THE element and STS marker sY579 is located between the arms of the duplication. *Eag*I (E) and *Mlu*I (M) restriction sites are shown.

The resulting plasmid DNAs were electroporated (Gene Pulser; Bio-Rad) into DH10B *Escherichia coli* cells (Life Technology) at a setting of 25 μF, 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 *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 nonduplicated 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 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. In all three cases we replicated our findings using samples from multiple, unrelated men.

Representative results of hybridizing *DAZ* cosmids to interphase fibroblast nuclei are shown in Figs. 2A and 2B. Cosmid 18E8 (5′ *DAZ*) generated two signals in 75% of the 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 signals (28%), two signals (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 unrelated men. We consistently observed evidence that each cluster contained two *DAZ* genes in head-to-head orientation (Figs. 2D and 2E). 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.

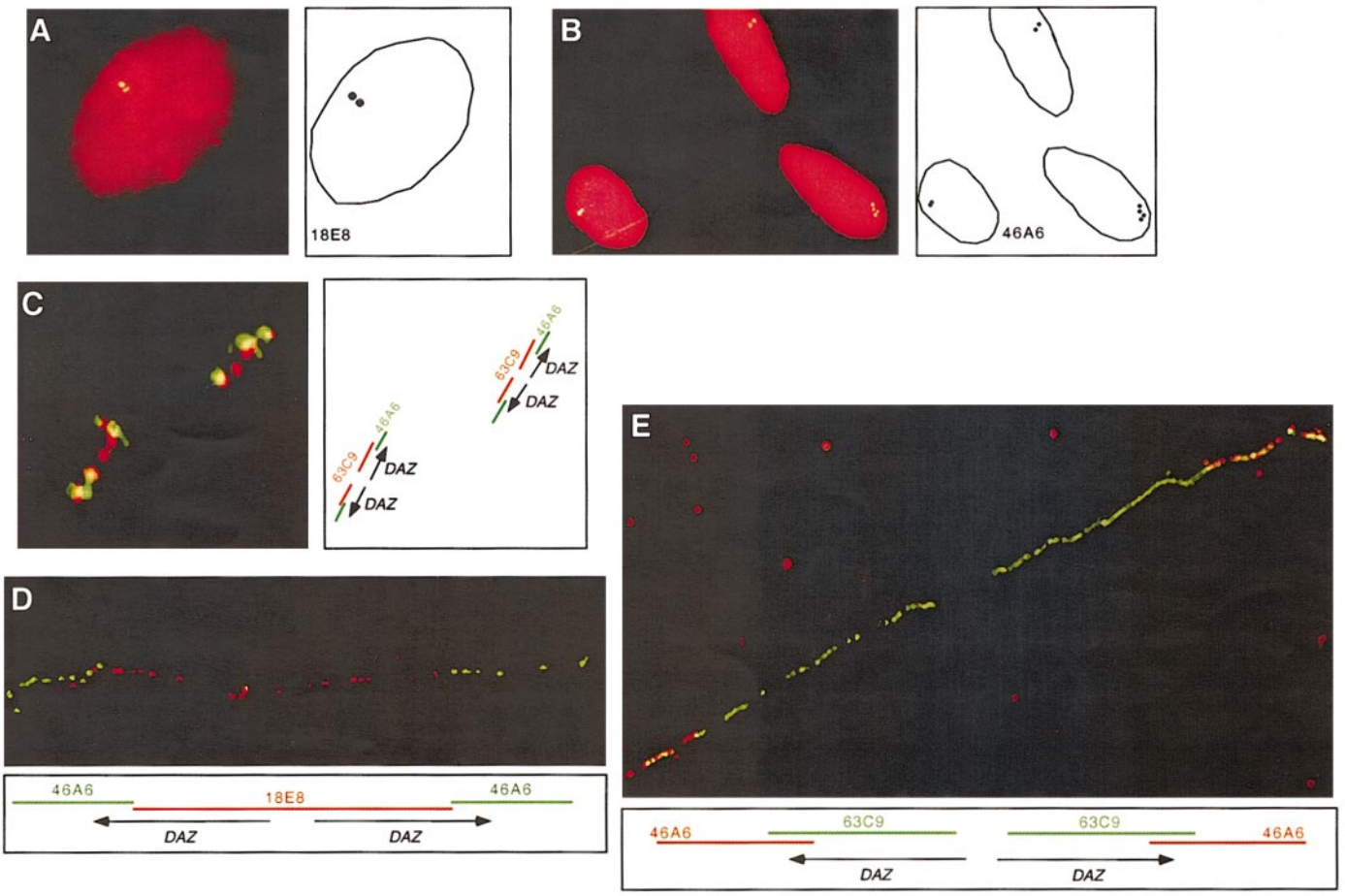

FIG. 2. FISH analysis of human Y chromosomes using *DAZ* cosmid probes. (**A**) Cosmid 18E8 (containing 5′ portions of two *DAZ* genes; Fig. 1) and (**B**) cosmid 46A6 (3′ portion of *DAZ*) on interphase nuclei from human fibroblasts. (**C**) Two-color FISH with cosmids 63C9 (containing all but exon 1; red) and 46A6 (green) on extended chromatin fibers from spermatozoa. (**D**) Cosmids 18E8 (red) and 46A6 (green) or (**E**) 63C9 (green) and 46A6 (red) in two-color FISH on extended chromatin fibers from lymphocytes.

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 four- to fivefold 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 single nucleotide variants that might be useful, we reasoned, in distinguishing among the *DAZ* genes. We developed PCR/restriction-digestion assays to type the BACs for these single nucleotide variants. Typing of the 16 BACs for three sequence variants (sY581/*Sau*3A, sY586/*Taq*I, and sY587/*Dra*I) 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—*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 the 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* was 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.

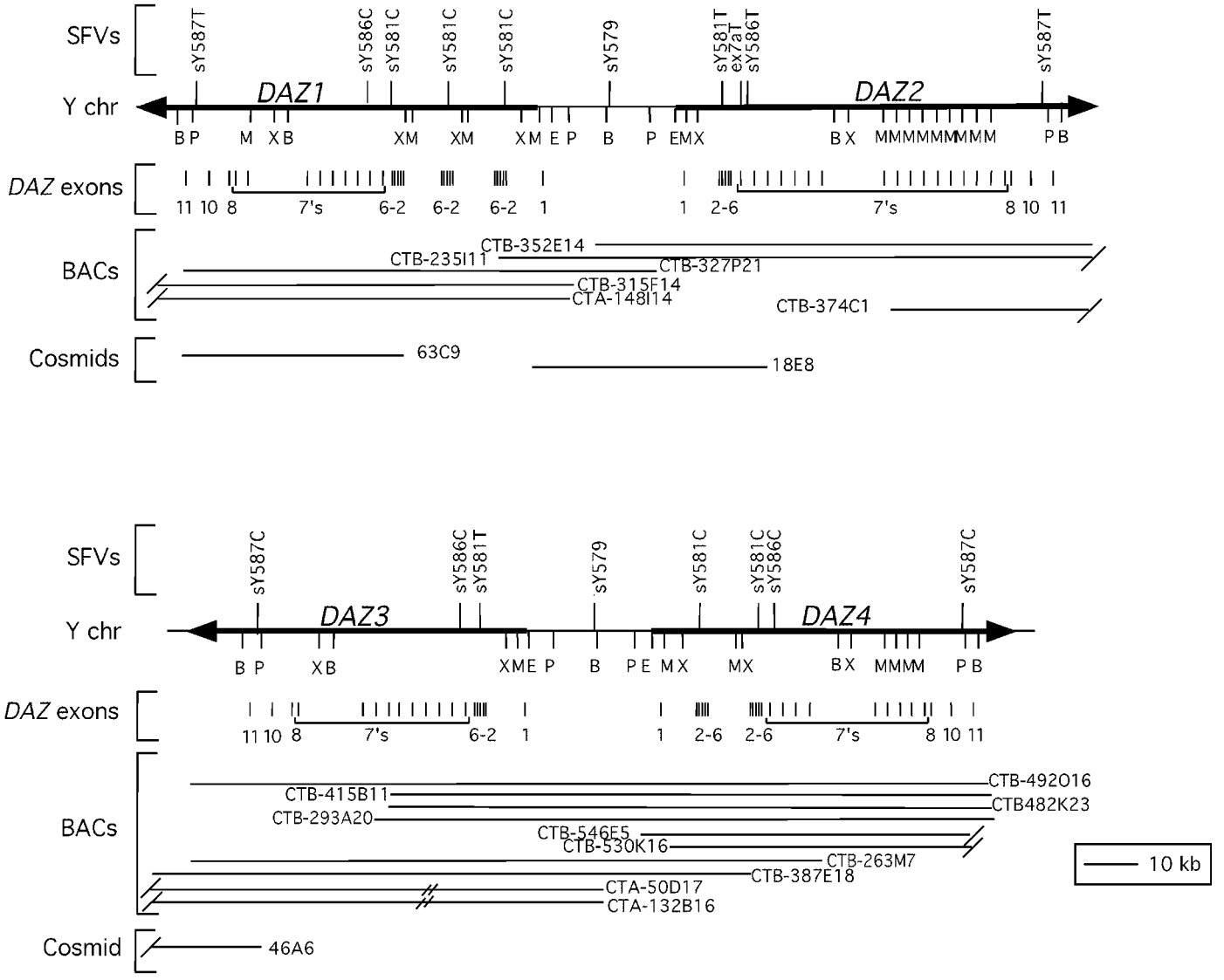

FIG. 3. Genomic organization of four *DAZ* genes in two clusters as inferred from analysis of BAC and cosmid clones. Oversized arrows indicate the direction of transcription of the *DAZ* genes; restriction sites: B, *Bam*HI; P, *Pme*I; M, *Mlu*I; X, *Xho*I; E, *Eag*I. Shown above the arrows are sequence family variants (SFVs; Table 1) that distinguish between *DAZ* genes; e.g., "sY586C" indicates that a C is present at the variable nucleotide position in sY586. The position of STS sY579 is also shown. Below the arrows, *DAZ* exons are numbered. BACs prefixed with CTA derive from a different male donor than BACs prefixed with CTB. In *DAZ2* BACs, the precise number of 2.4-kb repeats bearing an *Mlu*I site is not known; the number shown is our estimate. The small slashes on the *DAZ3*-containing CTA BACs (CTA-50D17 and CTA-132B16) indicate that they contain two more 2.4-kb repeats (two more copies of exon 7) than the *DAZ3*-containing CTB BACs.

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′-

TABLE 1

PCR/Restriction-Digestion Typing of Sequence Family Variants That Distinguish between *DAZ* Genes

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	130, 59, 63	189, 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

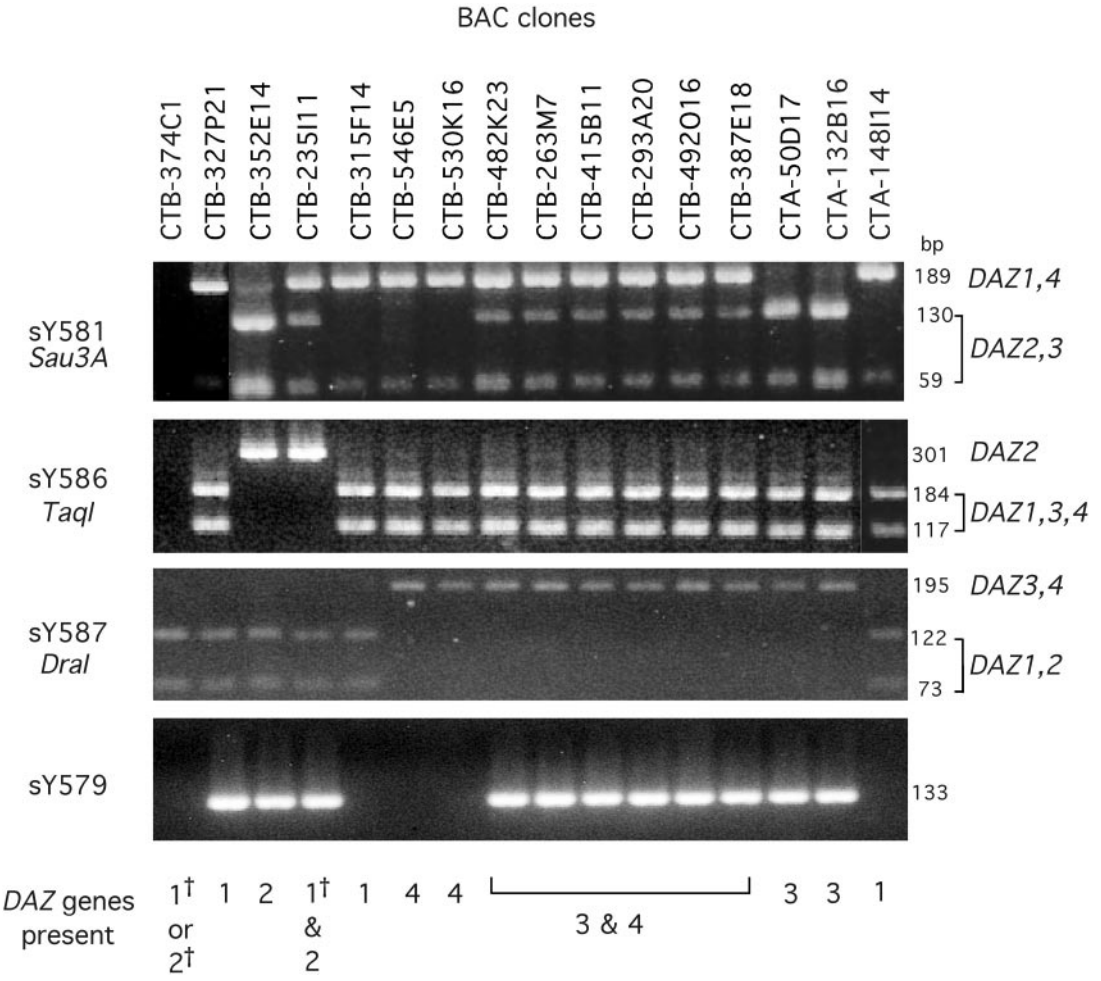

FIG. 4. Sequence family variants in *DAZ* BAC clones scored by PCR/restriction-digestion analysis. Assays listed along the left are described in Table 1; positions of SFVs within *DAZ* genes are shown in Fig. 3. Along the right are listed fragment sizes (in bp) and *DAZ* genes giving rise to each fragment. sY579 maps between the 5′ ends of inverted *DAZ* genes (Fig. 1). Listed at the bottom of each lane is the *DAZ* gene(s) present in that BAC clone; †Only a portion of the indicated *DAZ* gene is present in that BAC.

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 a 10.8-kb Unit within DAZ1 and DAZ4

We then compared the four *DAZ* genes at the 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 *Pme*I, 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*, 50 kb; and *DAZ4*, 55 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

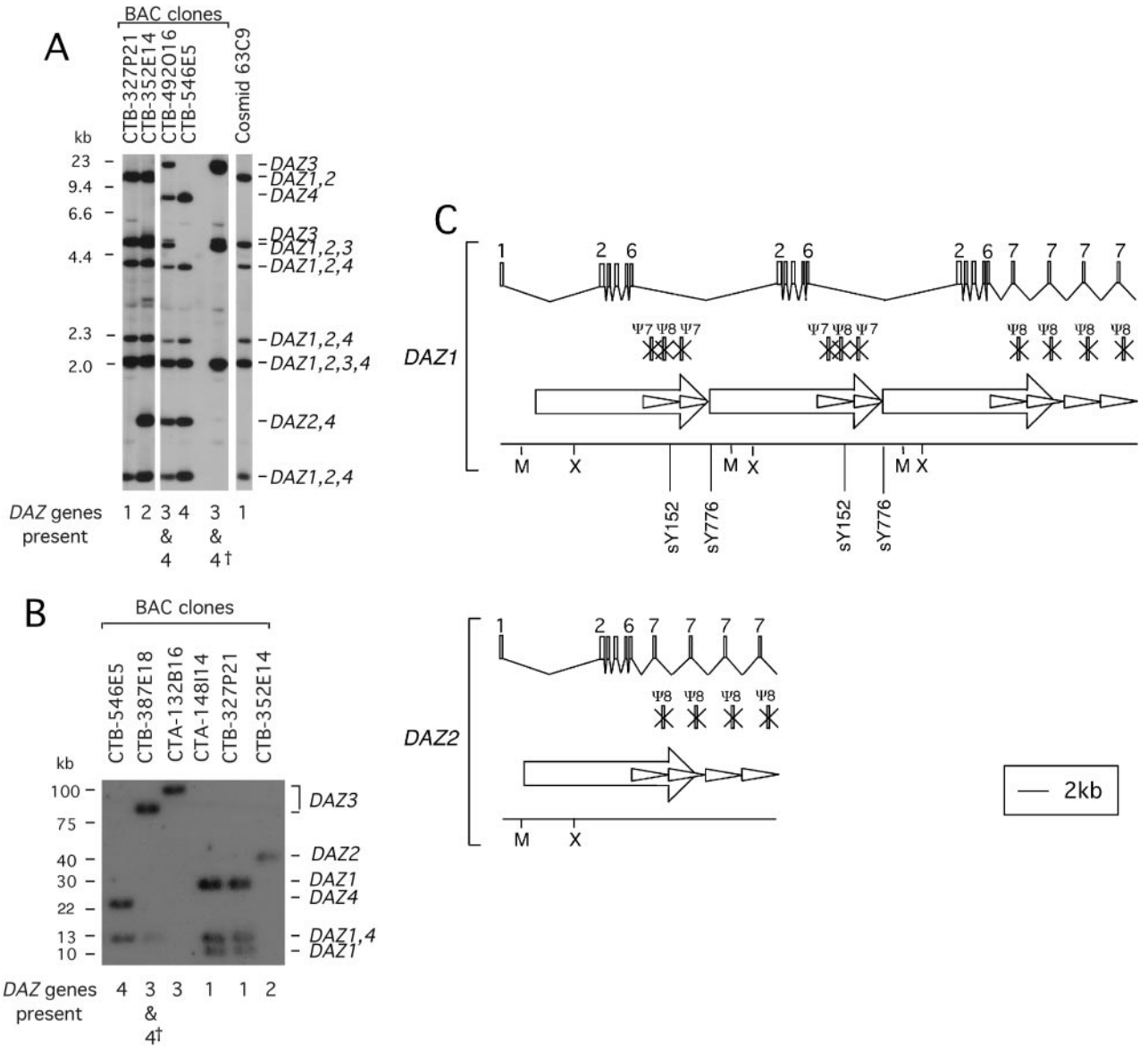

FIG. 5. Tandem repeats of 2.4- and 10.8-kb units in *DAZ* genes. (**A**) Southern hybridization of 2.4-kb repeat probe pDP1649 to *Taq*I-digested *DAZ* BAC and cosmid DNAs. Along the right are listed the *DAZ* genes giving rise to each fragment. Listed at the bottom of each lane is the *DAZ* gene(s) present in that BAC or cosmid clone; †only a portion of the indicated *DAZ* gene is present. Like cosmid 63C9, which has eight tandem 2.4-kb repeats interrupted by a LINE element (Saxena *et al.*, 1996), all *DAZ*-containing BACs display multiple large hybridizing fragments. (**B**) Southern hybridization of a PCR fragment spanning *DAZ* exons 2 and 3 to *Mlu*I-digested *DAZ* BAC DNAs. (**C**) Schematic diagram of 5′ portions of *DAZ1* and *DAZ2* genes, with, respectively, three tandem copies or one copy of the 10.8-kb repeat (large open arrow). The 2.4-kb repeats are shown as smaller open arrowheads. Exons and pseudoexons (ψ) are indicated above the repeats; restriction sites (M, *Mlu*I; X, *Xho*I) and positions of STS markers sY152 and sY776 (which detects the junction between tandem 10.8-kb repeats) are shown at the bottom.

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 *DAZ4*, 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 pulsed-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 *DAZ3* (BAC CTA-132B16). However, we observed three or two hybridizing *Mlu*I fragments in BACs containing *DAZ1* (BACs CTA-148I14 and CTB-327P21) or *DAZ4* (BAC CTB-546E5), respectively. These results suggested that *DAZ1* and *DAZ4* contained, respectively, three and two copies of exons 2 and 3. Additional Southern blot studies of BAC DNAs revealed that exons 4–6 are also

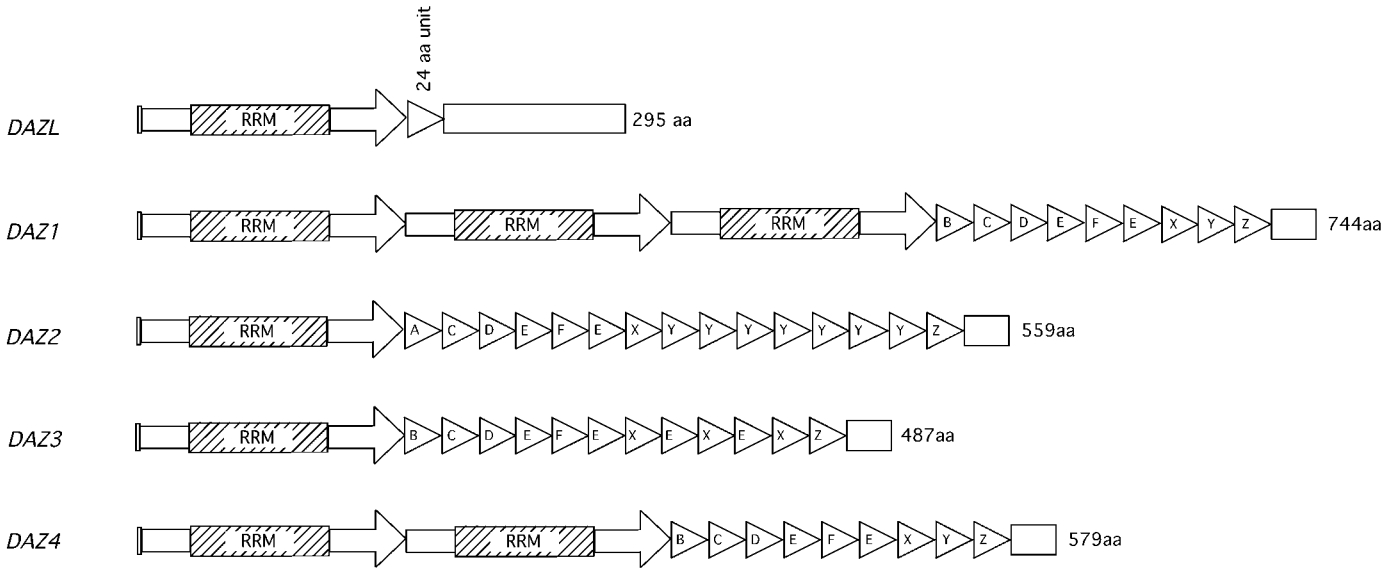

FIG. 6. Predicted human DAZL (autosomal) and DAZ (Y-linked) proteins. Large arrow: 165-amino-acid unit encompassing 82-amino-acid RRM. Smaller arrowheads: 24-amino-acid units labeled according to the nomenclature of Yen *et al.* (1997). The C-terminal portion (open rectangle) of the DAZL protein has no similarity to the C-terminal portions of the DAZ proteins.

present three times in *DAZ1* and twice in *DAZ4* (data not shown). These findings were corroborated and extended by nucleotide sequence analysis of *DAZ1* BAC CTA-148I14 (GenBank Accession No. AF164343; D. Schwartz and colleagues, Madison, pers. comm. Aug. 1999), which revealed a tandem triplication of a 10.8-kb genomic unit encompassing exons 2 through 6.

In 1992, Vollrath *et al.* (1992) identified sY152, an STS mapping to deletion interval 6D. This STS was subsequently shown to be deleted in infertile men with *AZFc*/*DAZ* deletions (Reijo *et al.,* 1995). However, Kent-First *et al.* (1999) 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. We find, as depicted in Fig. 5C, that sY152 is located within the 10.8-kb repeats in the *DAZ1* and *DAZ4* 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 *DAZ3,* and the third to *DAZ4* 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 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 *DAZ3* 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 *DAZ4*, but not in *DAZ3* (Fig. 3). [Note the absence of *Mlu*I sites among the tandemly repeated exon 7s of *DAZ3* (Fig. 3).] The *DAZ3* 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 *DAZ3*.

The third full-length cDNA clone (pDP1680/pDP1681) most likely derives from *DAZ4*, but we cannot exclude the possibility that it derives from *DAZ1*. This cDNA clone differs dramatically from the *DAZ2* and *DAZ3* 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 this 495-nucleotide unit within the cDNA corresponds well to the tandem duplication of the 10.8-kb unit in the *DAZ4* genomic locus (Fig. 3). The putative *DAZ4* 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 *DAZ4* or possibly from *DAZ1*. Neither cDNA extended sufficiently 5′ to include exon 1, but both appeared to derive from transcripts in which exons 2 through 6 were (at least) duplicated, consistent with their being derived from either *DAZ1* or *DAZ4*. Both cDNAs contained nine copies of exon 7 (BCDEFEXYZ), as also found in putative *DAZ4* cDNA pDP1681, suggesting that these clones may be derived from *DAZ4*.

Partial sequence analysis of the remaining 12 *DAZ* cDNA clones revealed no additional classes of transcripts. Nonetheless, the existence of polymorphic or alternatively spliced forms cannot be ruled out.

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 reinterpreted 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 the 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 *et al.* (1997; Yen, 1998). Through Southern blot analysis of genomic DNAs, Yen *et al.* (1997) determined that most Y chromosomes carry at least three *DAZ* genes. 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 (1998) constructed a YAC-based, long-range restriction map on which she has located three *DAZ* genes. 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.

Recently, seven *DAZ* BACs that originate from a different male donor were partially sequenced as part of an ongoing collaboration between our laboratory and the Washington University Genome Sequencing Center (Table 2). The emerging draft sequence appears to be consistent with the existence of four *DAZ* genes arranged as shown in Fig. 3.

Sequence Family Variants

Distinguishing among and unambiguously identifying each of the four *DAZ* genes were technically challenging tasks. We recognized eventually that *DAZ1*, *DAZ2*, *DAZ3*, and *DAZ4* possessed different numbers of intragenic (2.4- 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- and the 10.8-kb repeat arrays were far too

TABLE 2

***DAZ* Genes Present in Sequenced BAC Clones in GenBank**

DAZ BAC clone	GenBank Accession No.	*DAZ* genes present	sY579	sY581[a]	sY586	sY587	sY776
CTA-148I14	AF164343	*DAZ1*	−	C	C	T	+
RP11-290O3	AC010089	*DAZ3/4*[b]	+	T	C	C	−
RP11-263A15	AC007039	*DAZ2/1*[b]	+	TC	T	T	+
RP11-539D10	AC006338	*DAZ4*[b]	−	T	C	C	+
RP11-226N11	AC006990	*DAZ2*	+	T	T	T	−
RP11-26D12	AC006982	*DAZ3/4*	+	T	C	C	+
RP11-70G12	AC006983	*DAZ2*[b]	−	−	−	T	−
RP11-140H23	AC053490	*DAZ1/2*[b]	+	TC	TC	T	+

[a] sY581 does not distinguish between *DAZ3* and *DAZ4* in the man from whom the RP11 BAC library was constructed. Refer to Fig. 3 for map positions of variant nucleotides assayed by SFV-STSs sY581, sY586, and sY587.
[b] Partial *DAZ* gene.

large to allow PCR amplification across them (as one might do in the case of mini- or microsatellites). Second, the DNA sequences of the 10.8-kb repeats appear to be identical to one 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.

In the end, we identified individual *DAZ* genes primarily based on subtle sequence differences—especially basepair 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 (for example, single nucleotide variation or 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 the structural and functional analyses 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 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 *DAZ4* 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 *DAZ4*. Finally, the array of exon 7 repeats that is predicted to occur in *DAZ1* transcripts is very similar to that observed in *DAZ4*. 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 DAZ4) differ functionally from DAZ proteins with one RRM (as predicted for DAZ2 and DAZ3).

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 3 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

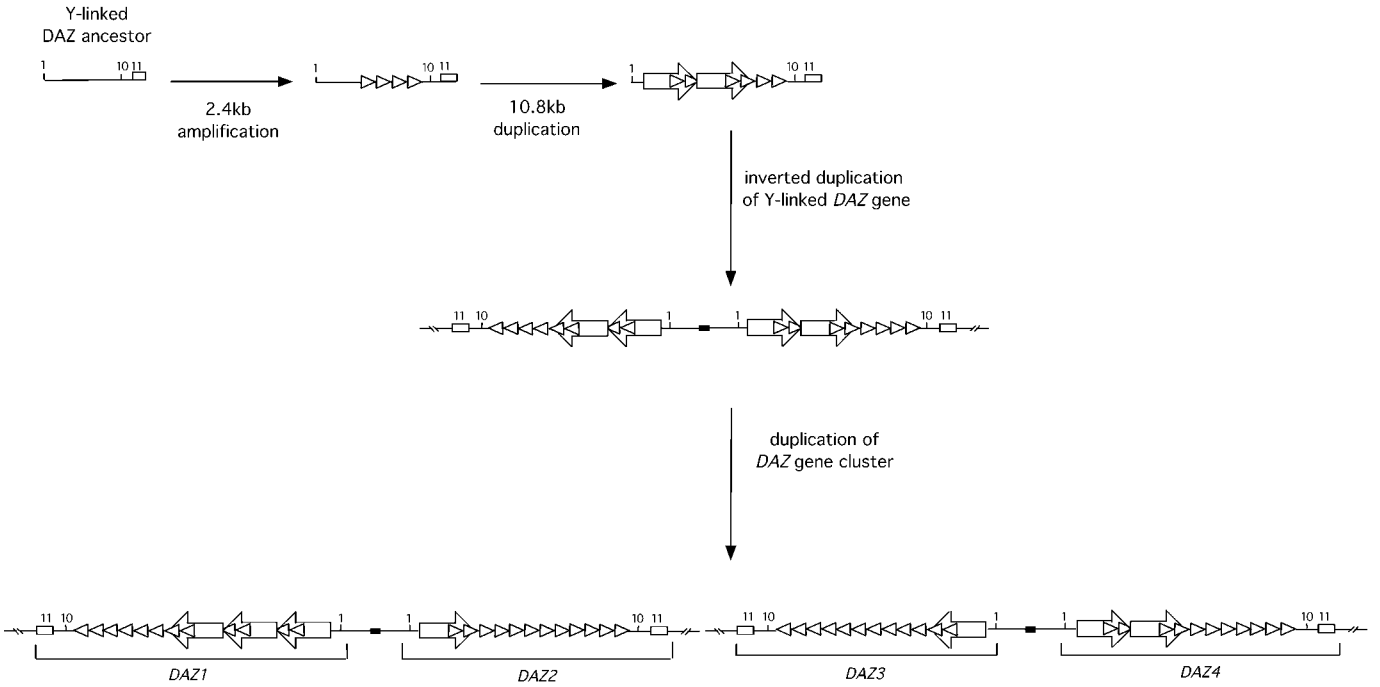

FIG. 7. An evolutionary model to account for four *DAZ* genes in two clusters on the human Y chromosome. Following transposition to the Y chromosome, the ancestral *DAZ* gene underwent amplification of the 2.4- and 10.8-kb units and pruning of many exons (see text). The precise order of events is speculative; other orders cannot be excluded. Exons 1,10, and 11 of each gene are shown. The 2.4-kb repeat unit is represented by the small arrowhead; the 10.8-kb repeat unit is represented by the larger open arrow. The THE element between inverted *DAZ* genes is shown as small black box. The relative orientation of the *DAZ1/DAZ2* and *DAZ3/DAZ4* clusters is not known.

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 *DAZ4* 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.

TABLE 3

Description of Repeats in *DAZ* Genes

		3′ repeat	
	5′ repeat	All but penultimate copy	Penultimate copy
Size of genomic unit	10.8 kb	2.4 kb	2.4 kb
Exons included	2, 3, 4, 5, 6, ψ7, ψ8, ψ7	7, ψ8	7, 8
Size of cDNA unit	495 bp	72 bp	107 bp
Number of amino acids encoded	165 (including RRM)	24	35+ frameshift

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 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 *DAZ4* 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. This research was 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:** 551–563.

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:** 2017–2024.

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.

Hull, M. G. R., Glazener, C. M. A., Kelly, N. J., Conway, D. I., Foster, P. A., Hinton, R. A., Coulson, C., Lambert, P. A., Watt E. M., and Desai, K. M. (1985). Population study of causes, treatment and outcome of infertility. *British Medical Journal* **291:** 1693–1697.

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 human chromosome 11. *Genomics* **21:** 538–550.

Reijo, R., Lee, T. Y., Salo, P., Alagappan, R., Brown, L. G., Rosenberg, M., Rozen, S., Jaffe, T., Strauss, D., Hovatta, O., de la Chapelle, A., 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, P. H., 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. (1998). A long-range restriction map of deletion interval 6 of the human Y chromosome: A region frequently deleted in azoospermic males. *Genomics* **54:** 5–12.

Yen, 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.