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.

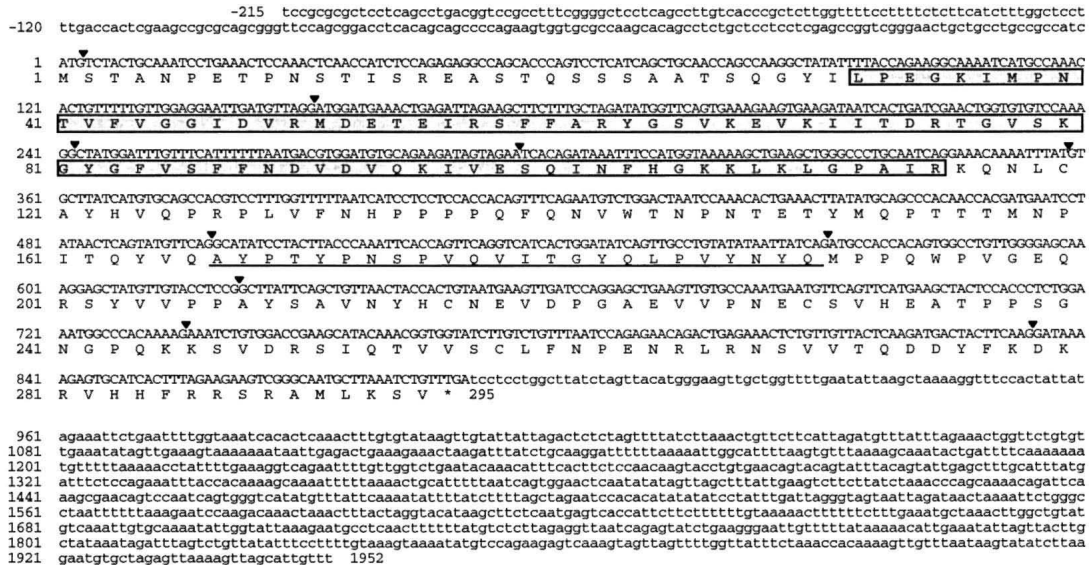

Fig. 1 Human *DAZH* cDNA sequence (clone pDP1648) and predicted amino acid sequence of encoded protein. RNP/RRM domain of protein is boxed. The single 24-amino-acid 'DAZ repeat' is underlined. Arrowheads above nucleotide sequence depict probable locations of ten introns (inferred by homology to *DAZ*; 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

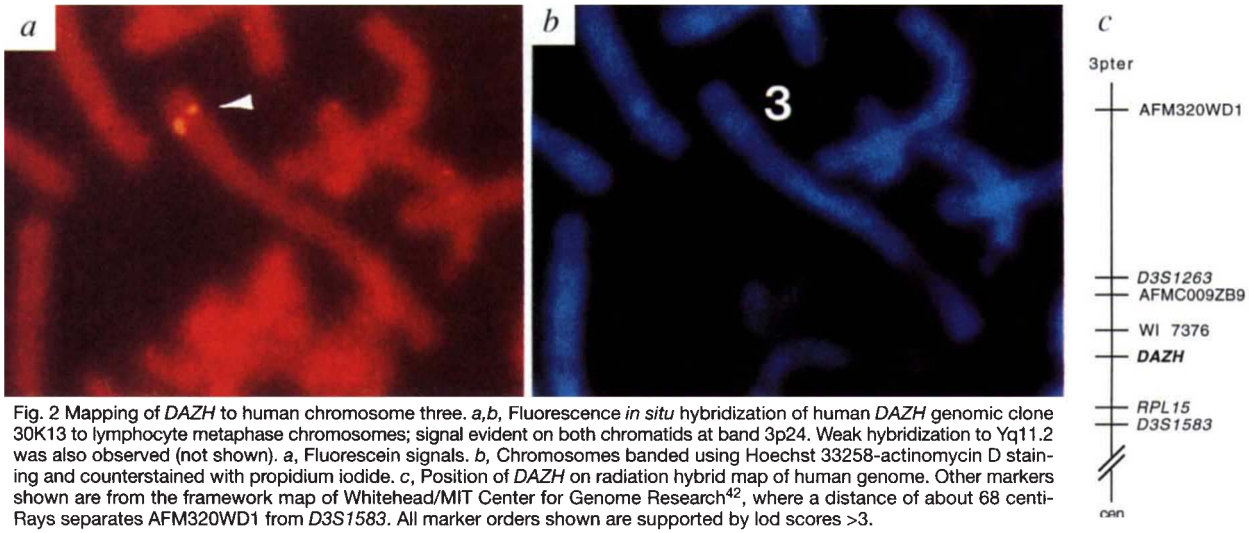

Fig. 2 Mapping of *DAZH* to human chromosome three. *a,b,* Fluorescence *in situ* hybridization of human *DAZH* genomic clone 30K13 to lymphocyte metaphase chromosomes; signal evident on both chromatids at band 3p24. Weak hybridization to Yq11.2 was also observed (not shown). *a*, Fluorescein signals. *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.

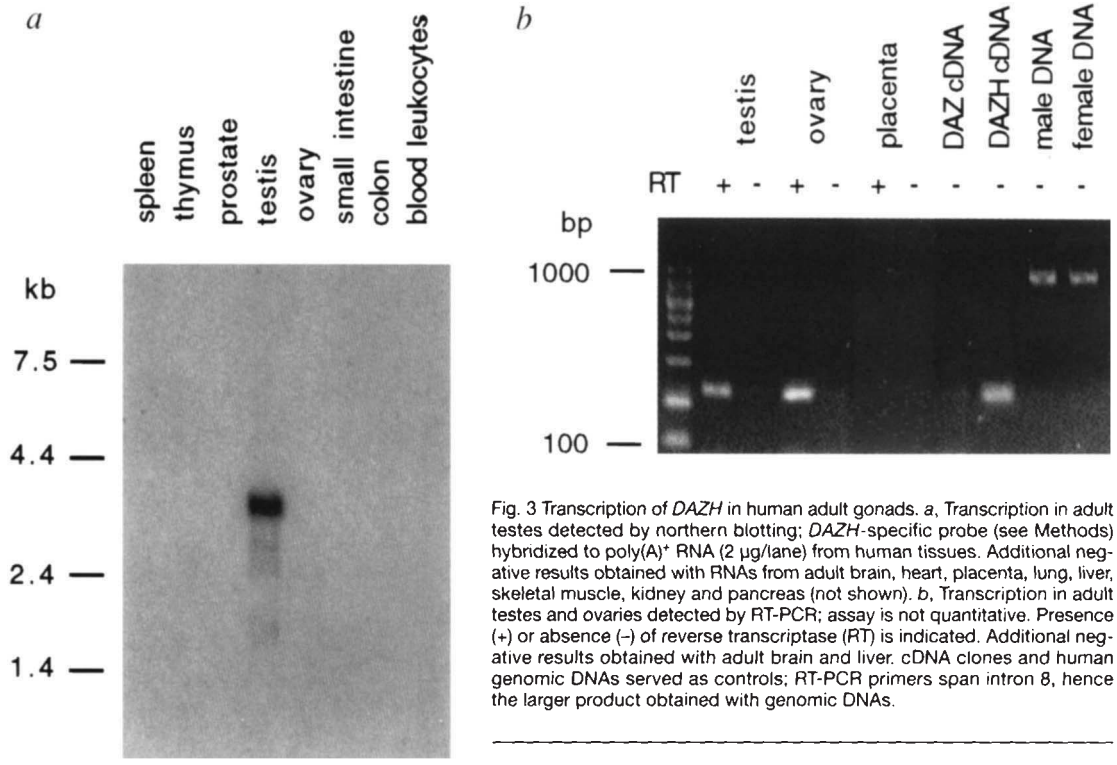

Fig. 3 Transcription of *DAZH* in human adult gonads. *a*, Transcription in adult testes detected by northern blotting; *DAZH*-specific probe (see Methods) hybridized to poly(A)$^+$ RNA (2 μg/lane) from human tissues. Additional negative results obtained with RNAs from adult brain, heart, placenta, lung, 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.

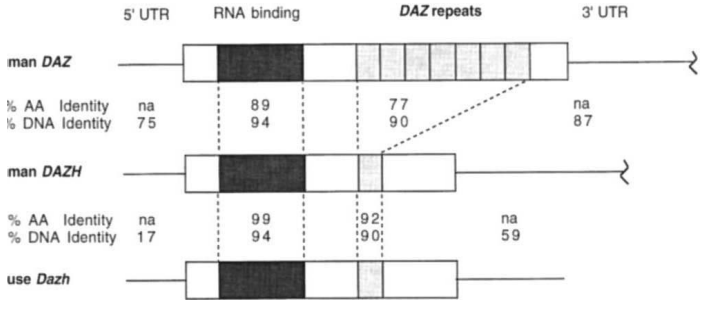

Fig. 4 Comparison of human *DAZ*, human *DAZH*, and mouse *Dazh* transcripts and encoded proteins[19,23,24]. This gene family encodes proteins with a single RNA-binding domain of the RRM/RNP type[43,44]. The human and mouse DAZH proteins have one copy of a 24-amino-acid unit that is tandemly repeated 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.

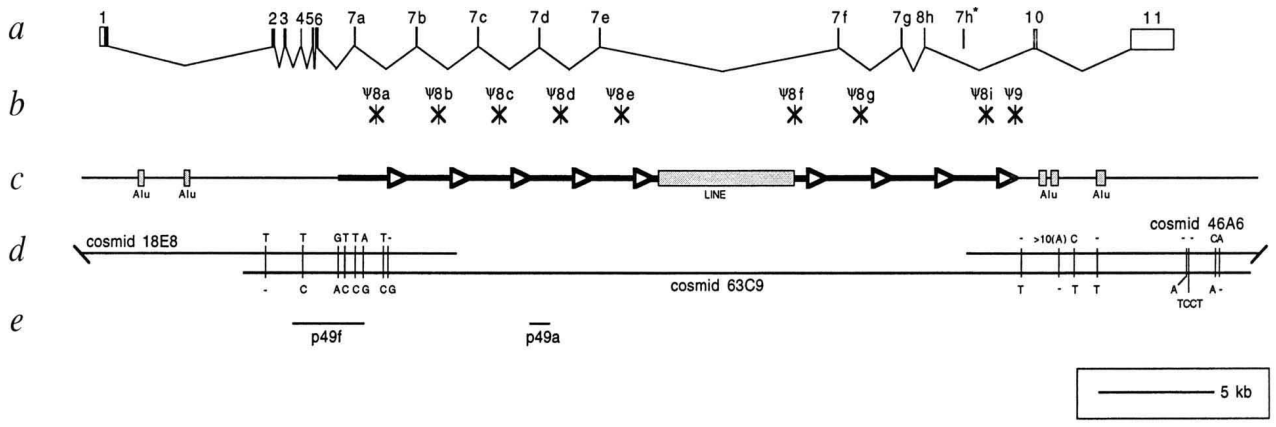

Fig. 5 Schematic representation of genomic DNA sequence from *DAZ* gene cluster on human Y chromosome. *a*, *DAZ* transcription unit. Exons numbered according to scheme outlined in Fig. 7; coding regions in black; UTRs in white. 7h, potentially an exon, has not been identified in sequenced cDNA clones (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,

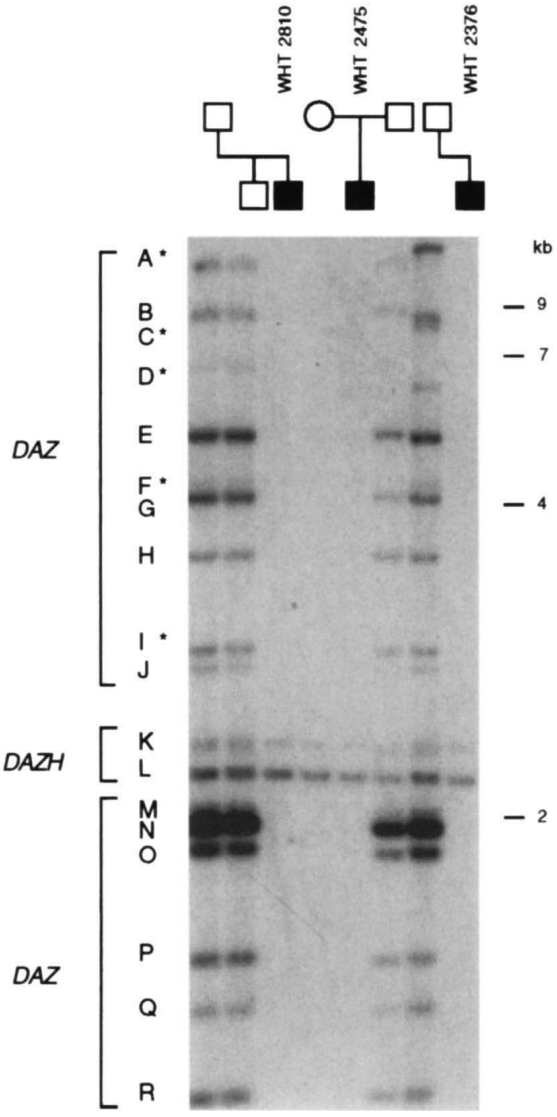

Fig. 6 *DYS1* probe p49f hybridized to Southern blot of *Taq*I-digested genomic DNAs (5 μg/lane) from three azoospermic men with *de novo* deletions of *AZF* region (and their immediate relatives; ref. 19 and R.A., Robert Oates, D.C.P, unpublished results). By convention[26], *Taq*I fragments hybridizing 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

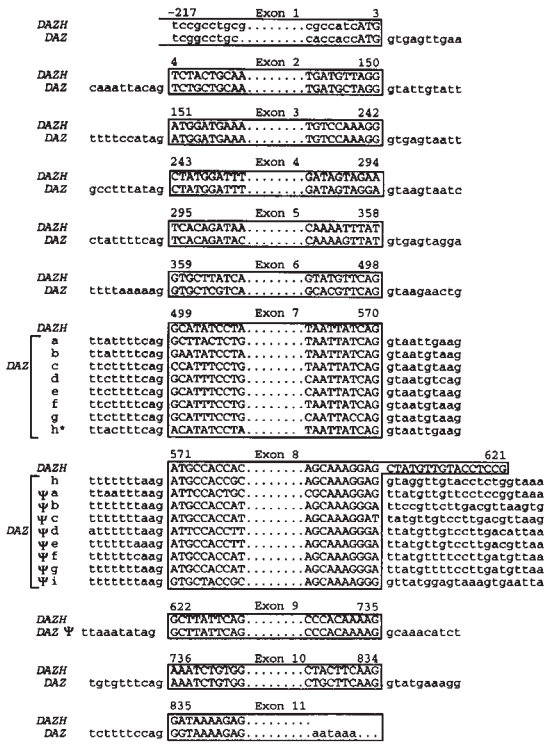

Fig. 7 Exons and pseudoexons of the human *DAZ* and *DAZH* genes. The figure is arranged in 11 tiers corresponding to the 11 exons of *DAZH* (boundaries inferred by homology to *DAZ*). In each tier the *DAZH* exon is shown in the top line, and below are shown all homologous regions, both exons and pseudoexons, 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 57 °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).