

## A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region

Sjoerd Repping,<sup>a,b</sup> Saskia K.M. van Daalen,<sup>b</sup> Cindy M. Korver,<sup>b</sup> Laura G. Brown,<sup>a</sup>  
Janet D. Marszalek,<sup>a</sup> Judith Gianotten,<sup>b</sup> Robert D. Oates,<sup>c</sup> Sherman Silber,<sup>d</sup>  
Fulco van der Veen,<sup>b</sup> David C. Page,<sup>a</sup> and Steve Rozen<sup>a,\*</sup>

<sup>a</sup>Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA

<sup>b</sup>Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Academic Medical Center, Amsterdam, The Netherlands

<sup>c</sup>Department of Urology, Boston University Medical Center, Boston, MA 02118, USA

<sup>d</sup>Infertility Center of St. Louis, St. Luke's Hospital, St. Louis, MO 63017, USA

Received 31 October 2003; accepted 12 December 2003

Available online 4 February 2004

### Abstract

The human Y chromosome is replete with amplicons—very large, nearly identical repeats—which render it susceptible to interstitial deletions that often cause spermatogenic failure. Here we describe a recurrent, 1.8-Mb deletion that removes half of the azoospermia factor c (*AZFc*) region, including 12 members of eight testis-specific gene families. We show that this “b2/b3” deletion arose at least four times in human history—likely on inverted variants of the *AZFc* region that we find exist as common polymorphisms. We observed the b2/b3 deletion primarily in one family of closely related Y chromosomes—branch N in the Y-chromosome genealogy—in which all chromosomes carried the deletion. This branch is known to be widely distributed in northern Eurasia, accounts for the majority of Y chromosomes in some populations, and appears to be several thousand years old. The population-genetic success of the b2/b3 deletion is surprising, (i) because it removes half of *AZFc* and (ii) because the gr/gr deletion, which removes a similar set of testis-specific genes, predisposes to spermatogenic failure. Our present findings suggest either that the b2/b3 deletion has at most a modest effect on fitness or that, within branch N, its effect has been counterbalanced by another genetic, possibly Y-linked, factor.

© 2004 Elsevier Inc. All rights reserved.

**Keywords:** Y chromosome, human; Chromosomal deletion; Inversion (genetics); Gene deletion; Infertility, male; Polymorphism; Haplotypes; Gene duplication; Evolution, molecular

The *AZFc* region of the human Y chromosome is critical for normal spermatogenesis, and the recurrent b2/b4 deletion, which removes all of *AZFc*, results in spermatogenic failure [1–3]. Evidence for deletions of part of *AZFc* was reported previously, although the precise nature and size of these deletions were not determined [4–7]. More recently, two deletions of part of *AZFc* were predicted, molecularly characterized, and shown to result from ectopic homologous recombination between amplicons [8,9]. One of these deletions, the gr/gr deletion, is a risk factor for spermatogenic failure, while the other, the b1/b3 deletion, is so rare that its effects on spermatogenesis have not been assessed. Because

of the importance of *AZFc* to spermatogenesis and the possibility that still other deletions of part of *AZFc* might exist, we searched for such deletions using *AZFc*-specific STS markers (Fig. 1A).

### Results

We screened 1563 men and found 25 who lacked sY1191 but possessed all flanking STSs (Fig. 1A and Tables 1 and 2). This STS pattern differs from previously described deletions that affect *AZFc*, including the b1/b3 and gr/gr deletions [3,9,10]. We investigated the nature of the new deletion in light of the observation that homologous recombination between amplicons causes nearly all

\* Corresponding author. Fax: (617) 258-5578.

E-mail address: [rozen@wi.mit.edu](mailto:rozen@wi.mit.edu) (S. Rozen).

previously described interstitial Y-chromosomal deletions involving *AZFc* [3,9,10]. Inspection of the reference sequence of *AZFc* shows that a single homologous recombination event between amplicons cannot produce a

chromosome lacking only *sY1191* (Fig. 1A). However, two mutational pathways, each consisting of an inversion followed by a deletion, could generate this STS pattern (Fig. 1). One pathway consists of an inversion involving

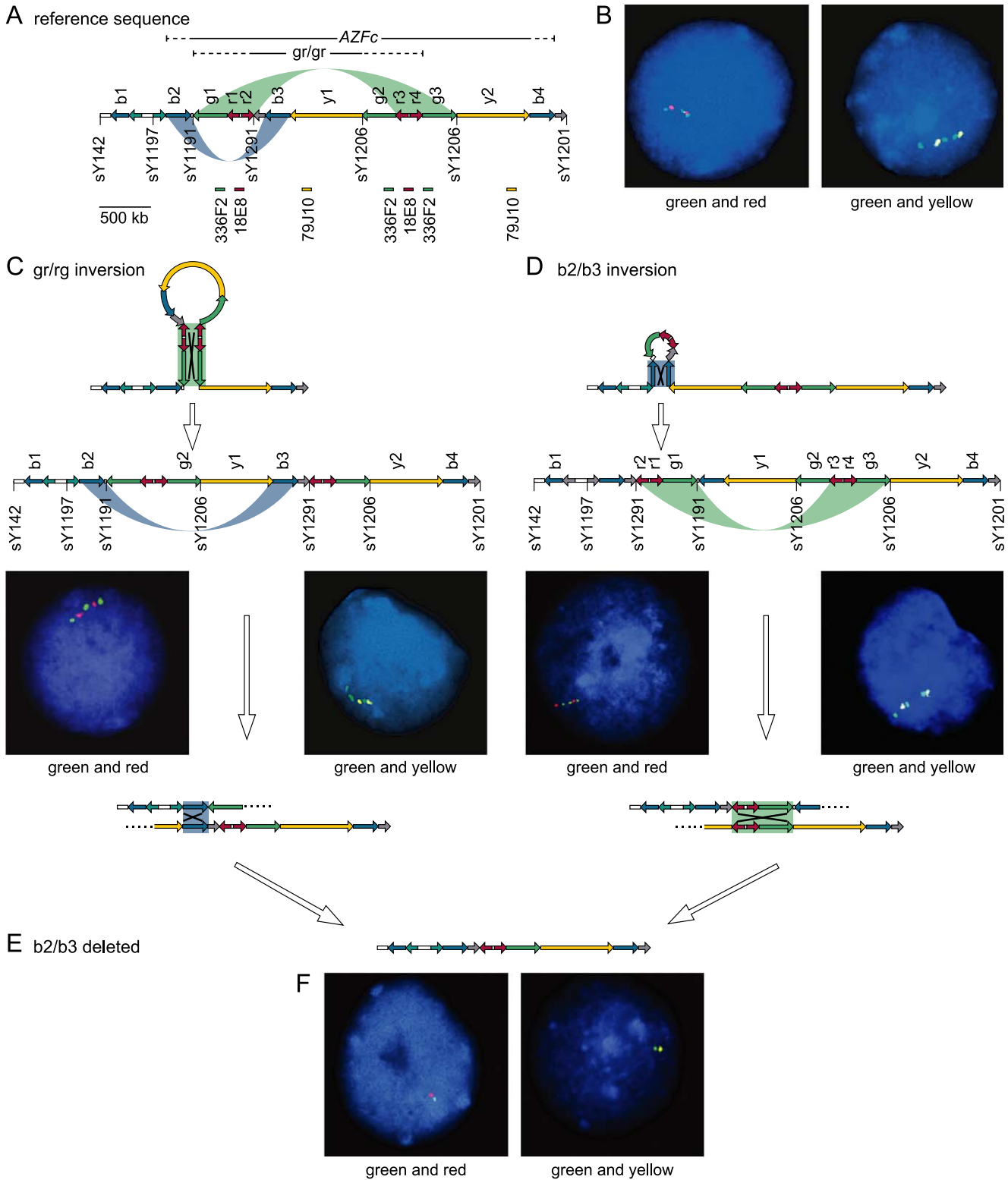


Fig. 1.

Table 1  
Men screened for deletions in *AZFc*

Category	Number of men tested	Number of men with sY1191-minus-only genotype
Spermatogenic failure <sup>a</sup>	1089	7
Normal spermatogenesis <sup>a</sup>	148	2
Unknown spermatogenesis <sup>a</sup>	326	16
Total	1563	25

<sup>a</sup> See Materials and methods.

the green and red amplicons (a “gr/rg” inversion) followed by a deletion between amplicons b2 and b3 (Fig. 1C). The other pathway consists of an inversion between amplicons b2 and b3, followed by a deletion involving the red and green amplicons (an “rg/rg” deletion; Fig. 1D). Both pathways produce the same final arrangement of amplicons (Fig. 1E).

We then used FISH to determine whether the sY1191-deleted chromosomes had this predicted arrangement of amplicons. Of 14 chromosomes tested, we found 11 with this arrangement (Table 2 and Figs. 1E and 1F), while 3 seemed to have undergone a duplication after the deletion (Table 2 and Fig. 2). These duplications are reminiscent of the duplications observed subsequent to the gr/gr deletion [9].

If either of the two hypothesized inversion–deletion pathways caused the sY1191 deletions, then there might be chromosomes with an inversion only, as the inversion and deletion events would not likely occur in the same generation. We used FISH to search for these inversions in a panel of Y chromosomes that represented the full breadth of the Y chromosome’s genealogical tree. Specifically, we examined one chromosome from each of the 44 branches in Fig. 3. We detected both of the predicted inversions at reasonably high frequencies: 6/44 for the gr/rg inversion (Figs. 1C and 3) and 3/44 for the b2/b3 inversion (Figs. 1D and 3). These data do not indicate whether one of the two mutational pathways predominates in generation of chromosomes lacking only sY1191. However, we termed the final arrangement of amplicons

the “b2/b3” deletion, referring to the targets of homologous recombination in the more common gr/rg inversion (Fig. 1C).

We investigated the dynamics of b2/b3 deletions over the course of human history by taking advantage of the Y chromosome’s clonal transmission and knowledge of its genealogical tree [11–13]. We found that 21 of the 25 b2/b3 deletions cluster in a single family of Y chromosomes, branch N [13] (Table 2 and Fig. 3). Furthermore, all branch N chromosomes were b2/b3 deleted.

These findings shed new light on a previously described polymorphism in branch N. Y chromosomes in branch N were found to lack a specific 6-kb band in *EcoRI* Southern blots probed with 50f2 [14]. Our electronic analysis shows that 50f2 should hybridize to a 6-kb *EcoRI* fragment that is 16 kb distal to sY1191 and that would be removed in the b2/b3 deletion (Fig. 1). Thus, the b2/b3 deletion appears to explain the absence of this fragment among branch N chromosomes.

The b2/b3-deleted chromosomes outside of branch N were distributed over three other Y haplotypes, all of which contained many undeleted chromosomes (Fig. 3 and Table 2). The rarity of b2/b3-deleted chromosomes outside of branch N (only 4/1542) and their apparent universality within branch N suggest that b2/b3-deleted chromosomes in this branch descended from a single deleted founder. We conclude that the b2/b3 deletion arose independently once in branch N and at least once in each of the other three haplotypes.

How old and widespread are b2/b3-deleted branch N chromosomes? Branch N contains chromosomes with the derived allele at the Tat polymorphism [13], an allele whose age and distribution have been widely studied [15,16]. This allele, then, offers lower bounds on the age and distribution of branch N chromosomes. It is widespread in northern Eurasia and extremely common in some populations, including Eastern Siberian Yakuts (85%), Finns (60%), and Estonians (47%) [15,16]. Indeed, the geographical distribution of Tat-derived chromosomes and their microsatellite diversity suggest that this polymorphism is several thousand

Fig. 1. Inversion–deletion models of the origins of b2/b3 deletions. (A) The ampliconic complex embedding *AZFc*, shown to scale [3]. The central bar depicts the organization of the constituent amplicons, which are color-coded; sequences with the same color are >99.9% identical. The genomic extents of the b2/b4 *AZFc* deletion and the gr/gr deletion are shown on top. STSs used to detect the b2/b3 deletion are indicated immediately below the central bar. Locations detected by FISH probes RP11-336F2 (green [22]), cosmid 18E8 (red [18]), and RP11-79J10 (yellow [22]) are also shown. The green and blue arches indicate the regions involved in the first homologous recombination events in the two pathways. (B) FISH probes hybridized to interphase nuclei from a man (PD223) whose STS results are consistent with the reference sequence in (A). Hybridization with the green and red probes produced the expected pattern: green–red–green–red–green (left). Hybridization with the green and yellow probes also produced the expected pattern: green–yellow–green–green–yellow (right). (C) Pathway 1: gr/rg inversion (g1, r1, r2 recombining with r3, r4, g3; green-shaded box) followed by a b2/b3 deletion (blue-shaded box), both via homologous recombination. The blue arch in the inverted organization indicates the sequences involved in the deletion. sY1191 and one copy of sY1206 lie in the deleted region, accounting for the sY1191-minus-only genotype. Interphase nuclei (from PD116) hybridized as above showed the patterns expected for a gr/rg inversion: green and red probes produced the same pattern as in B, and green and yellow probes produced a pattern different from that of B: green–green–yellow–green–yellow. (D) Pathway 2: b2/b3 inversion (blue-shaded box) followed by rg/rg deletion (green-shaded box) both via homologous recombination. The green arch in the inverted organization indicates the sequences involved in the deletion. As in C, sY1191 and one copy of sY1206 lie in the deleted region. Interphase nuclei (from PD264) hybridized as above showed the patterns expected for b2/b3 inversion: green and red probes produced a pattern different from those of B and C: red–green–green–red–green, and green and yellow probes produced the same pattern as in B. (E) Final arrangement in both pathways. (F) Interphase nuclei of a man (WHT3420) with the STS signature indicating a b2/b3 deletion, hybridized with FISH probes as above. The patterns are as predicted: red–green (left) and green–yellow (right).

Table 2  
FISH analysis of men lacking sY1191 and their location in the genealogical tree of the Y chromosome

Sample ID	FISH		Branch of Y genealogy <sup>b</sup>	Spermatogenic phenotype
	Number of dots with red probe <sup>a</sup>	Organization of dots after cohybridization with green and red probes <sup>a</sup>		
<i>Men for whom FISH showed the b2/b3 deletion</i>				
AMC0135	1	G-R	F*(xHK)	Abnormal
AMC0172	1	G-R	I	Normal
PD321	1	G-R	N*(xN3)	Unknown
AMC0010	1	G-R	N3	Normal
PD427	1	G-R	N3	Unknown
WHT0716	1	G-R	N3	Unknown
WHT3420	1	G-R	N3	Abnormal
WHT4704	1	G-R	N3	Unknown
WHT4830	1	G-R	N3	Unknown
PD024	1	G-R	Q3	Unknown
PD066	1	G-R	Q3	Unknown
<i>Men for whom FISH indicated duplication subsequent to the b2/b3 deletion</i>				
AMC0171	2 <sup>c</sup>	Insufficient cells	N*(xN3)	Abnormal
PD403	2 <sup>c</sup>	G-R-G-R <sup>c</sup>	N*(xN3)	Unknown
WHT3645	3 <sup>c</sup>	R-G-R-R-G <sup>c</sup>	N*(xN3)	Abnormal
<i>Men for whom cells were unavailable</i>				
AMC0170			N3	Abnormal
AMC0173			N3	Abnormal
WHT4443			N3	Abnormal
WHT4869			N3	Unknown
WHT4895			N3	Unknown
WHT4906			N3	Unknown
YCC047			N3	Unknown
YCC048			N3	Unknown
YCC049			N3	Unknown
YCC050			N3	Unknown
YCC051			N3	Unknown

<sup>a</sup> See Fig. 1A.

<sup>b</sup> See Fig. 3.

<sup>c</sup> FISH result indicating duplication subsequent to b2/b3 deletion.

years old [15,16]. Thus, the b2/b3 deletion in branch N is at least as old.

## Discussion

The b2/b3 deletion removes 1.8 Mb of *AZFc*—a region essential for normal spermatogenesis—including 12 testes-specific genes or transcripts (Table 3); thus, the age, persistence, and geographical range of branch N Y chromosomes are surprising. Furthermore, a different deletion, the gr/gr deletion, has similar size and gene content (Fig. 1A and Table 3) and predisposes to spermatogenic failure [9].

It is not known whether the *AZFc* gene copies retained on b2/b3-deleted chromosomes are functionally equivalent to those retained on gr/gr-deleted chromosomes. There is little difference in the numbers of genes affected by the two deletions; the b2/b3 deletion removes one more copy of each of three gene families, but spares at least one copy of each family (Table 3). Based on the Y-chromosomal reference sequence, the sequences of the gene copies remaining

after the two deletions are essentially identical, except in the case of *DAZ* [17]. Different members of the *DAZ* gene family are known to have various intragenic duplications [17,18] and might as a consequence not be functionally equivalent. Furthermore, the near sequence identity of the other genes does not preclude expression differences that might be due to, for example, differences between the two deletions in long-range sequence organization.

While seeking to understand the success of the b2/b3 deletion in branch N in light of the effects of the gr/gr deletion on spermatogenesis, it is important to consider the differences in the frequencies of the two deletions. We previously reported finding the gr/gr deletion in 14 different Y haplotypes and at a relatively high frequency, 3.5%, among men with spermatogenic failure [9]. This allowed us to assess its effects on spermatogenesis in diverse genetic contexts. In contrast, we found the b2/b3 deletion in only four men outside of branch N. Consequently, currently available data do not allow us to assess whether the b2/b3 deletion, in and of itself, is a risk factor for spermatogenic failure.

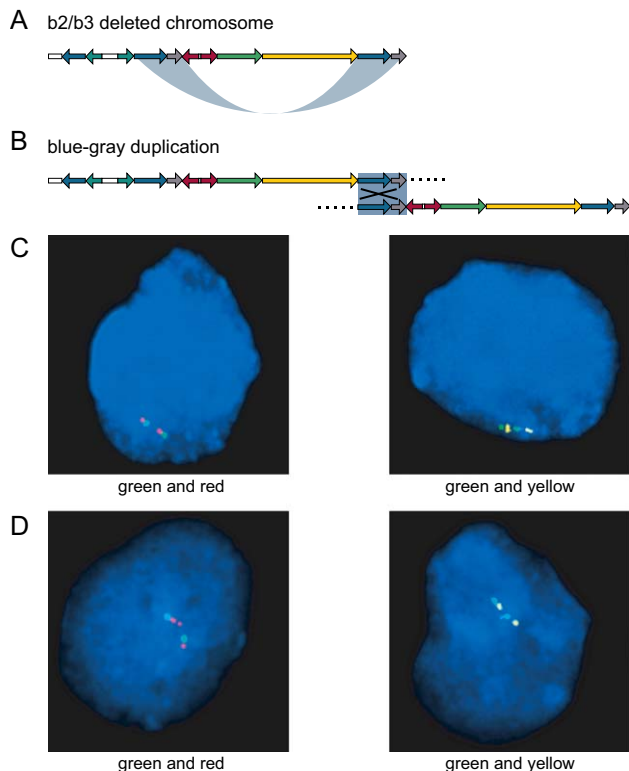


Fig. 2. Subsequent rearrangements of b2/b3-deleted chromosomes. (A) Organization of amplicons in a b2/b3-deleted chromosome, with the amplicons involved in the blue-gray duplication indicated by the arch. (B) Model of blue-gray duplication caused by homologous recombination (shaded box). (C) Interphase nuclei of a man (PD403) hybridized with probes as in Fig. 1. The results are as predicted for a blue-gray duplication: red-green-red-green (left) and green-yellow-green-yellow (right). (D) Interphase nuclei from WHT3645 hybridized with FISH probes as in Fig. 1. Homologous recombination between amplicons in a b2/b3-deleted chromosome cannot explain this result. However, the fact that this man's chromosome is in branch N and lacks only sY1191 suggests that it descended from a b2/b3-deleted chromosome. The organization indicated by the FISH results could be the result of nonhomologous rearrangement or of recombination between short repeats. Such mutations are much less common in the Y chromosome than homologous recombination between amplicons, but have been reported [3,10].

Indeed, we must consider the possibility that, within branch N, any reduction in fitness caused by the b2/b3 deletion might have been counterbalanced by a beneficial factor, perhaps located elsewhere in the Y chromosomes of branch N. In the absence of meiotic crossing over with a homologous chromosome, this factor would remain completely linked to the b2/b3 deletion. Alternatively, however, the b2/b3 deletion might have only a minimal effect on spermatogenesis and fitness. If so, small effective population sizes might have allowed branch N chromosomes to drift to high frequencies [19].

Investigating the possible effects of the b2/b3 deletion on spermatogenesis would likely require screening tens of thousands of men, because of (i) the possibility of a compensatory factor in branch N, (ii) the rarity of b2/b3 deletions outside of branch N, and (iii) the possibility that

any such effects may be weak. Studies of this scale are currently infeasible, but are among the goals of the Human Genome Project and may become possible in the future [20].

## Materials and methods

**Men studied.** Men with spermatogenic failure (Table 1) were patients from the Academic Medical Center and patients studied at the Whitehead Institute, all with idiopathic nonobstructive azoospermia (no sperm in semen) or idiopathic severe oligozoospermia (sperm count  $<10^7$ /ml or total sperm count  $<2 \times 10^7$ ). Men with normal spermatogenesis (Table 1) were patients from the Academic Medical Center with total sperm count  $>4 \times 10^7$  and normal sperm motility and morphology. Men with unknown spermatogenesis (Table 1) were included for analysis based on their Y haplotype without knowledge of their sperm count. Most of these samples were purchased from the NHGRI/NIGMS DNA Polymorphism Discovery Resource [21] (Coriell Cell Repositories). We prepared DNA from peripheral blood leukocytes or EBV-transformed lymphoblastoid cell lines.

**Deletion screening.** We carried out initial deletion screening using sY1191, followed by assaying the STSs in Fig. 1 on men lacking sY1191. Primers and PCR conditions are available from GenBank under the following accession numbers: sY142, G38345; sY1197, G67168; sY1191, G73809; sY1291, G72340; sY1206, G67171; and sY1201, G67170. We carried out two-color FISH as described previously [18].

**Y-chromosomal genealogy.** The Y genealogy used is based on those published in [11–13]. A subset of the men screened for deletions, including all men with b2/b3 deletions, was assigned to one of the branches of the genealogy by genotyping them at the Y polymorphisms shown (Fig. 3). References to the polymorphisms and primer pairs can be found in [13].

## Acknowledgments

We thank J. Alfoldi, J. Hughes, J. Koubova, P. Lombardi, S. Mastenbroek, H. Skaletsky, J. de Vries, and H. Westerveld for comments on the manuscript; M. Alamares, R.E. Berger, D.L. Bluestein, C. Bruning, C. Disteche, A.E. Donnenfeld, N.A. Ellis, S. Fallet, A. Garguilo, J.L. German, B.R. Gilbert, D. Gurwitz, M.F. Hamper, W.A. Hogge, P. Hoo, M. Jamehdor, T. Jenkins, K. Kepler, K. Monaghan, P. Patrizio, E. Pergament, J.C. Petrozza, K. Phelan, J. Popovic, B. Shapiro, M.C. Summers, U. Surti, J. Tollis, L. Weiss, and J. Weissenbach for samples; C. Tyler-Smith and T. Zerjal for assistance with LLY22g; and M.F. Hammer and P.A. Underhill for assistance with genealogical studies. This



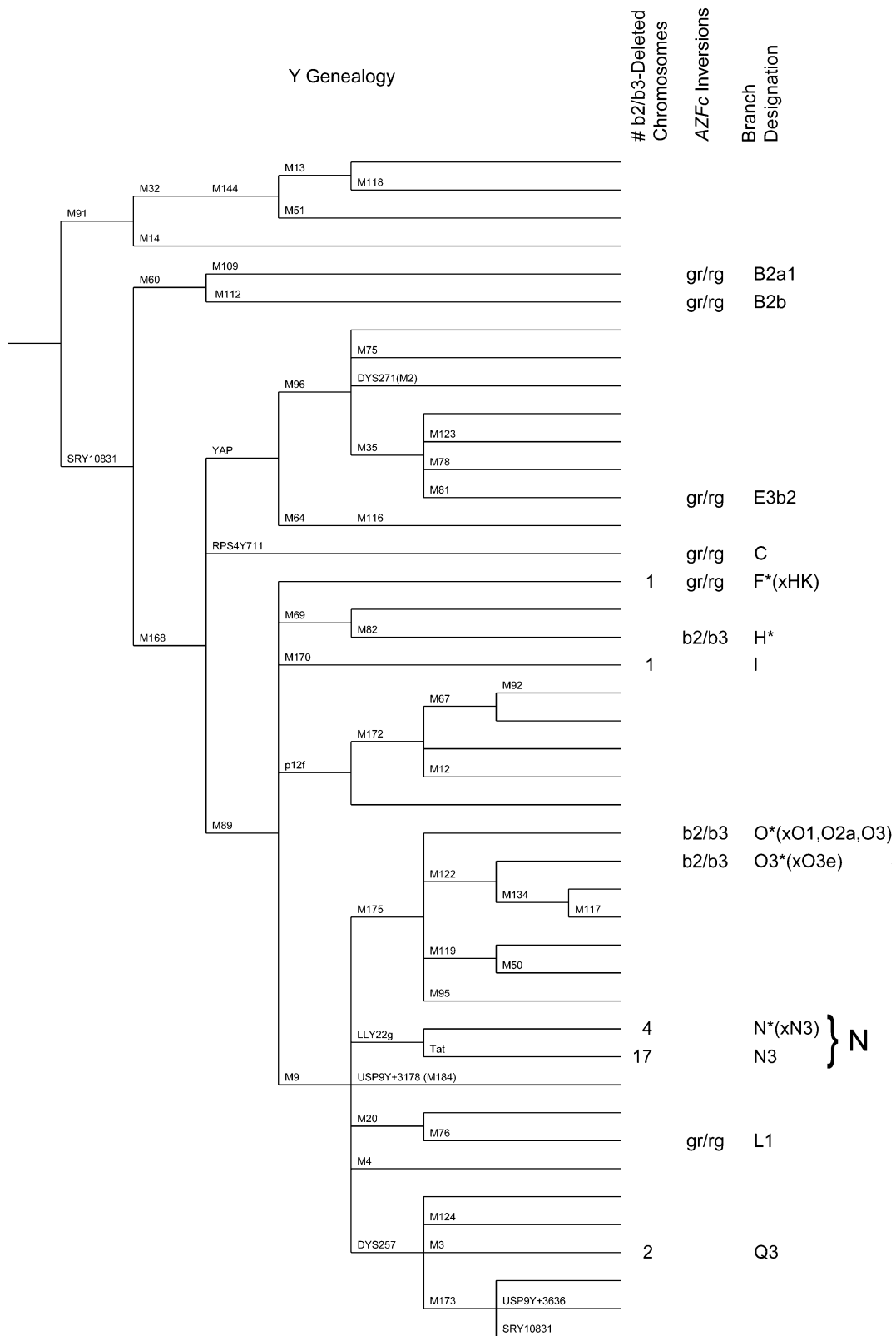


Fig. 3. Genealogical analysis of b2/b3 deletions and AZFc inversions. The genealogical tree of extant Y chromosomes is shown at the left [13]. The branches labeled with branch designations (rightmost column) are those in which we observed Y chromosomes with the b2/b3 deletion (Table 2) or one of the inversions predicted in Fig. 1. The 1563 men screened for the b2/b3 deletion included representatives of all branches shown in the genealogy. In branch N, 21/21 men were b2/b3 deleted; in branch F\*(xHK), 1/18; in branch I, 1/38; in branch Q3, 2/9.

Table 3

Effects of the b2/b3 deletion on gene and transcript families compared with other common deletions affecting *AZFc*

Gene or transcript family <sup>a</sup>	Number of copies present			
	Reference sequence (or gr/rg or b2/b3 inverted <sup>b</sup> )	b2/b3 deletion	gr/gr deletion <sup>c</sup>	b2/b4 ( <i>AZFc</i> ) deletion <sup>c</sup>
<i>RBMY</i>	6	6	6	6
<i>BPY2</i>	3	1	2	0
<i>DAZ</i>	4	2	2	0
<i>CDY1</i>	2	1	1	0
<i>PRY</i>	2	2	2	2
<i>CSPG4LY</i>	2	1	1	0
<i>GOLGA2LY</i>	2	1	1	0
<i>TTY3</i>	2	1	1	0
<i>TTY4</i>	3	1	2	0
<i>TTY5</i>	1	1	1	1
<i>TTY6</i>	2	2	2	2
<i>TTY17</i>	3	1	2	0
Total	32	20	23	11

<sup>a</sup> See [17] for precise gene locations and GenBank accession numbers.

<sup>b</sup> See Fig. 1.

<sup>c</sup> Adapted from [8].

work was supported by the National Institutes of Health, the Howard Hughes Medical Institute, and the Academic Medical Center.

## References

- [1] R. Reijo, et al., Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene, *Nat. Genet.* 10 (1995) 383–393.
- [2] P.H. Vogt, et al., Human Y chromosome azoospermia factors (*AZF*) mapped to different subregions in Yq11, *Hum. Mol. Genet.* 5 (1996) 933–943.
- [3] T. Kuroda-Kawaguchi, et al., The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men, *Nat. Genet.* 29 (2001) 279–286.
- [4] T. Bienvenu, C. Patrat, K. McElreavey, M. de Almeida, P. Jouannet, Reduction in the *DAZ* gene copy number in two infertile men with impaired spermatogenesis, *Ann. Genet.* 44 (2001) 125–128.
- [5] J.W.A. de Vries, et al., Reduced copy number of *DAZ* genes in sub-fertile and infertile men, *Fertil. Steril.* 77 (2002) 68–75.
- [6] S. Fernandes, et al., High frequency of *DAZ1/DAZ2* gene deletions in patients with severe oligozoospermia, *Mol. Hum. Reprod.* 8 (2002) 286–298.
- [7] S. Repping, et al., The use of spermHALO-FISH to determine *DAZ* gene copy number, *Mol. Hum. Reprod.* 9 (2003) 183–188.
- [8] P. Yen, The fragility of fertility, *Nat. Genet.* 29 (2001) 243–244.
- [9] S. Repping, et al., Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection, *Nat. Genet.* 35 (2003) 247–251.
- [10] S. Repping, et al., Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure, *Am. J. Hum. Genet.* 71 (2002) 906–922.
- [11] P.A. Underhill, et al., Y chromosome sequence variation and the history of human populations, *Nat. Genet.* 26 (2000) 358–361.
- [12] P.A. Underhill, et al., The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, *Ann. Hum. Genet.* 65 (2001) 43–62.
- [13] The Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Res.* 12 (2002) 339–348.
- [14] M.A. Jobling, et al., Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males, *Hum. Mol. Genet.* 5 (1996) 1767–1775.
- [15] T. Zerjal, et al., Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis, *Am. J. Hum. Genet.* 60 (1997) 1174–1183.
- [16] P. Lahermo, et al., Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami, *Eur. J. Hum. Genet.* 7 (1999) 447–458.
- [17] H. Skaletsky, et al., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes, *Nature* 423 (2003) 825–837.
- [18] R. Saxena, et al., Four *DAZ* genes in two clusters found in the *AZFc* region of the human Y chromosome, *Genomics* 67 (2000) 256–267.
- [19] T.M. Karafet, et al., High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter–gatherer way of life, *Hum. Biol.* 74 (2002) 761–789.
- [20] F.S. Collins, E.D. Green, A.E. Guttmacher, M.S. Guyer, A vision for the future of genomics research, *Nature* 422 (2003) 835–847.
- [21] F.S. Collins, L.D. Brooks, A. Chakravarti, A DNA polymorphism discovery resource for research on human genetic variation, *Genome Res.* 8 (1998) 1229–1231.
- [22] C. Tilford, et al., A physical map of the human Y chromosome, *Nature* 409 (2001) 943–945.