nature genetics

Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection

Sjoerd Repping[1,2], Helen Skaletsky[1], Laura Brown[1], Saskia K M van Daalen[2], Cindy M Korver[2], Tatyana Pyntikova[1], Tomoko Kuroda-Kawaguchi[1,3], Jan W A de Vries[2], Robert D Oates[4], Sherman Silber[5], Fulco van der Veen[2], David C Page[1] & Steve Rozen[1]

Many human Y-chromosomal deletions are thought to severely impair reproductive fitness, which precludes their transmission to the next generation and thus ensures their rarity in the population. Here we report a 1.6-Mb deletion that persists over generations and is sufficiently common to be considered a polymorphism. We hypothesized that this deletion might affect spermatogenesis because it removes almost half of the Y chromosome's *AZFc* region, a gene-rich segment that is critical for sperm production[1,2]. An association study established that this deletion, called gr/gr, is a significant risk factor for spermatogenic failure. The gr/gr deletion has far lower penetrance with respect to spermatogenic failure than previously characterized Y-chromosomal deletions; it is often transmitted from father to son. By studying the distribution of gr/gr-deleted chromosomes across the branches of the Y chromosome's genealogical tree, we determined that this deletion arose independently at least 14 times in human history. We suggest that the existence of this deletion as a polymorphism reflects a balance between haploid selection, which culls gr/gr-deleted Y chromosomes from the population, and homologous recombination, which continues to generate new gr/gr deletions.

Much of the human Y chromosome consists of long, Y-specific repeats called amplicons[1,3]. Homologous recombination between amplicons has been shown to generate deletions, commonly resulting in spermatogenic failure[1,4–7]. The *AZFc* region is comprised completely of amplicons (**Fig. 1a**) and is particularly susceptible to deletions. For example, the b2/b4 deletion, which spans 3.5 Mb and eliminates the entire *AZFc* region (**Fig. 1a**), is the most common known genetic cause of spermatogenic failure[1]. Inspection of the ampliconic structure of *AZFc* led to the prediction that two other deletions could arise there by homologous recombination[2]. Consistent with this prediction, we and others have previously reported evidence that some men have partial deletions of *AZFc*, although the precise nature and size of these deletions was not determined and their possible impact on spermatogenesis was unclear[8–12]. Here we report the identification of both predicted deletions[2]: the gr/gr deletion (1.6 Mb; **Fig. 1**) and the b1/b3 deletion (also 1.6 Mb; **Fig. 2**). Using sequence-tagged sites (STSs) to screen specifically for these deletions in 689 men, we found 22 apparent cases of the predicted gr/gr deletion and a single instance of the predicted b1/b3 deletion.

In light of the structural complexity and repetitive nature of the *AZFc* region, we sought to confirm these putative gr/gr and b1/b3 deletions using fluorescence *in situ* hybridization (FISH). The single b1/b3 deletion was confirmed (**Fig. 2**). Of the 22 gr/gr deletions, we were able to test 20 cases by FISH, and all were confirmed (**Fig. 1** and **Supplementary Table 1** online). In two of these 20 cases, the FISH studies identified a secondary duplication (**Fig. 3**).

It seemed likely that the gr/gr deletion would predispose men to spermatogenic failure, because it removes nine transcription units with testis-specific expression (**Table 1**) and part of *AZFc*, a region essential for normal spermatogenesis (**Fig. 1**). A reanalysis of the 689 men originally screened for the gr/gr deletion, which included 473 individuals known to have spermatogenic failure, provided preliminary support for this hypothesis. We sorted each of the 689 men into one of ten Y haplotypes to minimize the effects of stratification. In this sample, the gr/gr deletion was consistently more common among men with spermatogenic failure (**Supplementary Table 2** online).

Encouraged by these preliminary results, we then carried out a formal association study in a separate population. Here we compared the frequency of gr/gr deletions in men with spermatogenic failure to that in controls known to have normal spermatogenesis. Both the affected and unaffected men were drawn from the same clinic population, and

[1]Howard Hughes Medical Institute, Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [2]Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Academic Medical Center, Amsterdam, the Netherlands. [3]Reproduction Center, Tokyo Dental College, Ichikawa General Hospital, Ichikawa, Chiba, Japan. [4]Department of Urology, Boston University Medical Center, Boston, Massachusetts 02118, USA. [5]Infertility Center of St. Louis, St. Luke's Hospital, St. Louis, Missouri 63017, USA. Correspondence should be addressed to D.C.P. (page_admin@wi.mit.edu).

Published online 5 October 2003; doi:10.1038/ng1250

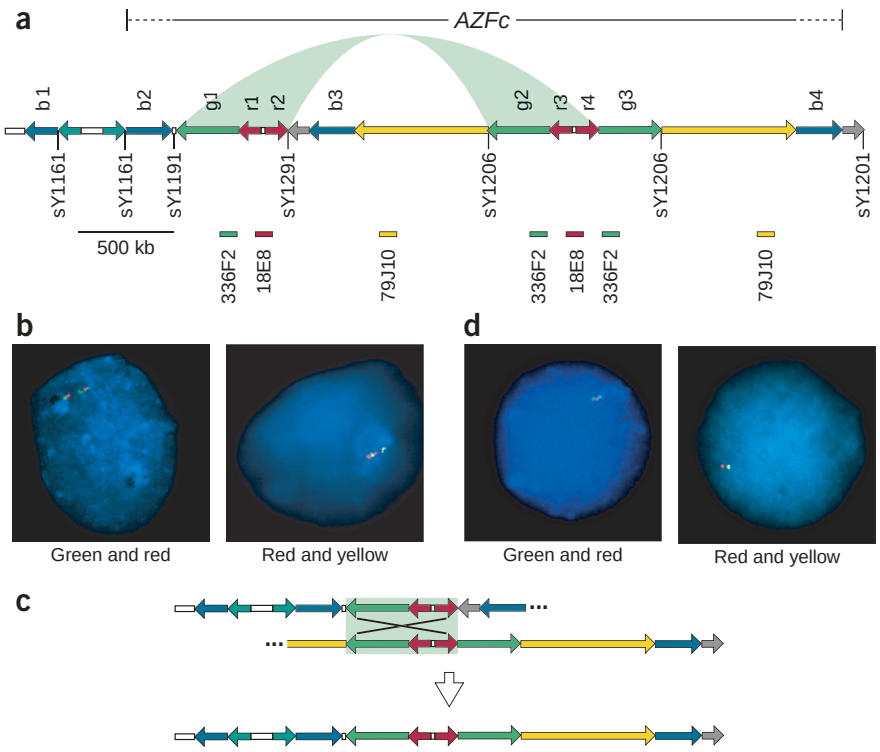

Figure 1 The gr/gr deletion. (**a**) The *AZFc* region of the Y chromosome[1]. The central bar depicts the organization of amplicons, including those labeled b1 through b4 (blue), g1 through g3 (green) and r1 through r4 (red). The green arch highlights the regions demarcating the gr/gr deletion (an abbreviation of Yen's designation g1/g2,r1/r3,r2/r4; ref. 2). Positions of STSs used to detect this deletion are indicated immediately below the central bar. Shown further below are hybridization sites for the following FISH probes: BAC RP11-336F2 (green, ref. 28), cosmid 18E8 (red, ref. 27) and BAC RP11-79J10 (yellow, ref. 28). At top is shown the extent of the recurrent, 3.5-Mb *AZFc* (b2/b4) deletion[1]. (**b**) FISH of interphase nuclei from a man (GM10470) in whom there is no gr/gr deletion. Hybridizations with red and green probes (left panel) and red and yellow probes (right panel) produced the expected patterns (green-red-green and red-yellow-red-yellow, respectively). (**c**) Simplest model of homologous recombination generating the gr/gr deletion. The green shaded box highlights the recombination targets. Recombination could be between sister chromatids or within a chromatid. Theoretically, more complex mutational pathways, each consisting of a series of homologous recombination events, could also generate the gr/gr deletion. The term 'gr/gr deletion' refers to the resulting organization of *AZFc* amplicons, not to any specific mutational pathway. (**d**) FISH of interphase nuclei from a gr/gr-deleted man (GM04535). Hybridizations produced the expected patterns (green-red-green and red-yellow).

the distribution of Y-chromosome haplotypes was similar in both groups (**Supplementary Table 3** online). This study identified a significant association between the gr/gr deletion and spermatogenic failure ($P < 0.014$; **Table 2**). As in our preliminary study, the prevalence of the gr/gr deletion was elevated consistently among men with spermatogenic failure (**Supplementary Table 4** online). This was true regardless of Y haplotype (as in our preliminary study), suggesting that the observed association is not due to stratification (**Supplementary Note** online). We conclude that the gr/gr deletion is a risk factor for spermatogenic failure.

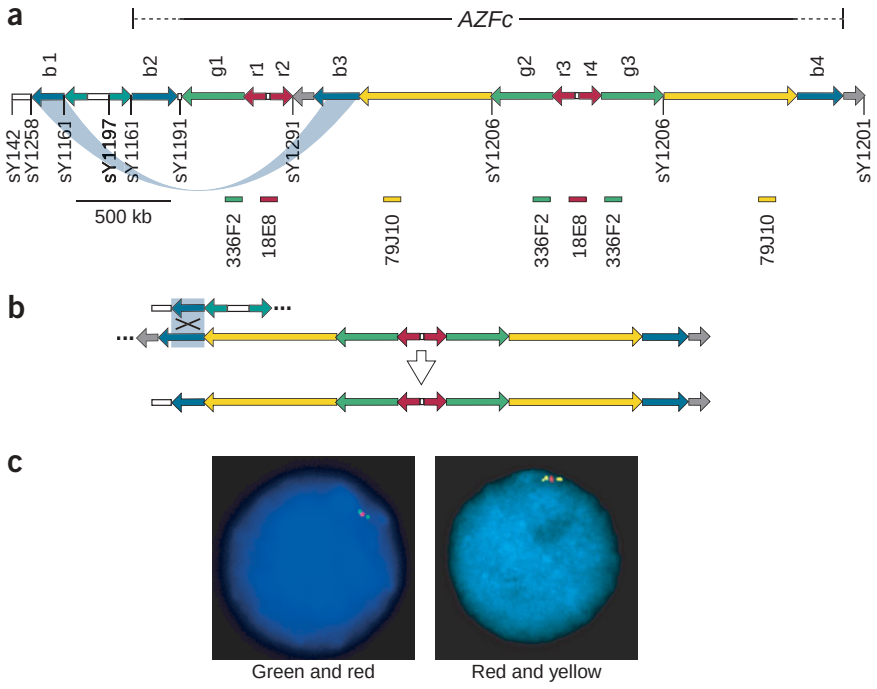

Figure 2 The b1/b3 deletion. (**a**) The *AZFc* region of the Y chromosome, as in **Figure 1a**. The blue arch highlights the b1 and b3 amplicons, which demarcate the deleted region. Positions of STSs used to detect this deletion are indicated immediately below the central bar. Shown further below are hybridization sites for FISH probes, as in **Figure 1a**. (**b**) Model of homologous recombination generating the b1/b3 deletion. The blue shaded box highlights the recombination targets. Recombination could be between sister chromatids or within a chromatid. (**c**) FISH of interphase nuclei from a man (WHT3453) with STSs results suggesting a b1/b3 deletion: he lacked sY1161, sY1197, sY1191 and sY1291 but possessed sY142, sY1258, sY1206 and sY1201. Hybridizations produced the patterns expected for a b1/b3-deleted chromosome (green-red-green and yellow-red-yellow). (In the case of a gr/gr deletion, the red and yellow probes would instead have produced the pattern red-yellow, as in **Figure 1**.)

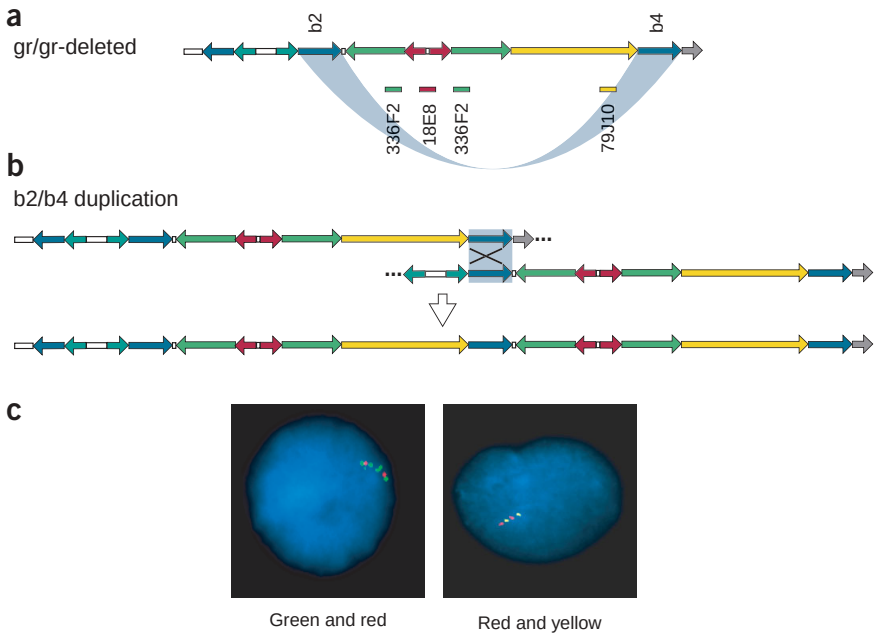

Figure 3 b2/b4 duplication arising in a gr/gr-deleted chromosome. (**a**) Organization of amplicons in a gr/gr-deleted chromosome. The blue arch highlights the b2 and b4 amplicons, which demarcate the duplicated segment. (**b**) Model of homologous recombination creating the b2/b4 duplication in a gr/gr-deleted chromosome. The blue shaded box highlights the recombination targets. This duplication is presumably the result of recombination between sister chromatids. (**c**) FISH of interphase nuclei from a man (WHT3173) with both a gr/gr deletion and a b2/b4 duplication. FISH probes are as in **Figure 1a**. Hybridization with green and red probes (left) produced the pattern green-red-green-green-red-green, and hybridization with red and yellow probes (right) produced the pattern red-yellow-red-yellow.

The gr/gr deletion does not completely eliminate any of the many testis-specific gene and transcription unit families on the Y chromosome. Instead, it reduces the copy number of eight such families (**Table 1**). We speculated that the dosage of one or more of these families affects the quantity of sperm produced. If so, then the secondary duplication observed in a few gr/gr-deleted chromosomes (**Fig. 3**) may act as a compensatory mutation by restoring gene copy number (**Table 1**).

Taking advantage of the clonal transmission of the Y chromosome (that is, without sexual recombination) and our knowledge of its genealogy[13–15], we examined the origins and dynamics of gr/gr deletions across the course of human history. We determined both the gr/gr status and the high-resolution Y haplotype of 368 men who collectively represent the full breadth of the chromosome's genealogical tree (43 branches; **Fig. 4**). We found no gr/gr-deleted chromosomes in 29 branches and found 13 branches in which these chromosomes were in the minority. Because 42 of 43 branches, including the deepest branches, consist partly or entirely of chromosomes that were not gr/gr-deleted, we concluded that the last common ancestor of modern human Y chromosomes was not gr/gr-deleted. Conversely, we found gr/gr-deleted chromosomes in 14 branches, including 1 that consisted entirely of these chromosomes. Examination of the tree indicates that the gr/gr deletion arose independently, one or more times, in each of these 14 branches.

One branch, D2b (**Fig. 4**), contained only gr/gr-deleted chromosomes. If the gr/gr deletion is a risk factor for spermatogenic failure, then men with D2b chromosomes should be at increased risk. Indeed, this probably accounts at a molecular level for an association between Y haplotype and spermatogenic failure observed in Japan, where D2b chromosomes are common. (They are rare in other populations, including the European and American populations that we studied[13].) Men with Y chromosomes roughly equivalent to branch D2b (**Supplementary Note** online) were found to be more likely than other Japanese men to have spermatogenic failure[16]. The gr/gr deletion that we have shown to be characteristic of branch D2b provides an explanation for this observation.

The present results have implications for the use of Y-chromosomal polymorphism as a tool for reconstructing the phylogeographic origins of modern human populations[13,14,17–19]. These reconstructions have routinely relied on the simplifying assumption that all Y polymorphisms are selectively neutral. This assumption should be reconsidered in light of the present evidence that a common Y-chromosomal variant can affect fertility and reproductive fitness.

Because the gr/gr deletion increases the risk of infertility, one might expect a substantial fraction of gr/gr deletions to have arisen *de novo*. Previously reported Y-chromosome deletions causing infertility, including deletions of the entire *AZFc* region, were almost always *de novo*[1,4,6,7,20–22]. In all four instances in which the father of an infertile, gr/gr-deleted man was available for testing, however, we found that the father's Y chromosome was also gr/gr-deleted. We conclude that the penetrance of the spermatogenic failure caused by the gr/gr deletion is much lower than that of previously characterized Y-chromosomal deletions.

Table 1 Genes and transcription units affected by deletions in the *AZFc* region

Gene or transcription unit	Numbers of copies present in: Reference Y chromosome	*AZFc* (b2/b4) -deleted	gr/gr-deleted	gr/gr-deleted with b2/b4 duplication	b1/b3-deleted
RBMY	6	6	6	6	4
BPY2	3	0	2	4	2
DAZ	4	0	2	4	2
CDY1	2	0	1	2	2
PRY	2	2	2	2	0
CSPG4LY	2	0	1	2	2
GOLGA2LY	2	0	1	2	2
TTTY3	2	0	1	2	2
TTTY4	3	0	2	4	2
TTTY5	1	1	1	1	0
TTTY6	2	2	2	2	0
TTTY17	3	0	2	4	2
Total	32	11	23	35	20

Modified from ref. 2.

Table 2 Numbers of gr/gr deletions among men with spermatogenic failure and among men with normal spermatogenesis

	gr/gr deletion	No gr/gr deletion	Total
Spermatogenic failure	9	237	246
Normal spermatogenesis	0	148	148
Total	9	385	394

$P < 0.014$ by Fisher's exact test, one-sided.

Because the gr/gr deletion is recurrent but tends to diminish fertility, we would expect the creation of new deletions to be balanced by their removal from the population by natural selection. Is there any evidence to suggest that this is the case? The targets for homologous recombination giving rise to gr/gr deletions are three times the size of the targets for the *AZFc* deletion (**Fig. 1**). Thus, one might suppose that new gr/gr deletions would be generated at a rate at least comparable to that of *AZFc* deletions. If so, and if the gr/gr deletion were selectively neutral, then population genetic theory suggests that >40% of Y chromosomes would be gr/gr-deleted (**Supplementary Note** online). In reality, across the diversity of the United States population (as sampled in the DNA Polymorphism Discovery Resource panel[23]), we observed only 2% of Y chromosomes (4 of 178 men tested) to be gr/gr deleted. This observed prevalence is consistent with the gr/gr deletion having a modest negative impact on reproductive fitness (**Supplementary Note** online).

In males, the X and Y chromosomes are present in only one copy, without homologs to complement gene deletions or other loss-of-function mutations. As a result, natural selection tends to remove new mutations from the population with a thoroughness and immediacy rarely seen on autosomes[24]. Thus, highly deleterious loss-of-function mutations on the sex chromosomes cannot become established as polymorphisms. In the case of X-linked hemophilia, for example, the frequency of mutated alleles is less than 0.02%, only a few times greater than the *de novo* mutation rate[25]. For *AZFc* and other Y-chromosome deletions causing spermatogenic failure with high penetrance, the prevalence of deleted Y chromosomes is again less than 0.03%, approximating the rate at which new deletions arise[1,7]. By contrast, the relatively high prevalence of the gr/gr deletion reflects a combination of low penetrance and high mutability, subject, of course, to random perturbation by genetic drift. To our knowledge, the gr/gr deletion is the first human polymorphism with strong evidence of a balance between haploid selection and new mutation.

METHODS

Affected individuals and controls. In the initial screen for gr/gr and b1/b3 deletions, we studied two groups: men with spermatogenic failure (as we suspected there could be an enrichment for the deletions among these men) and men selected to represent the worldwide diversity of Y-chromosomal haplotypes (as it was possible that some Y haplotypes could be enriched for these deletions). The group with spermatogenic failure consisted of men with nonobstructive azoospermia or severe oligozoospermia ($<5 \times 10^6$ sperm per ml semen), all from couples presenting with infertility at Boston University Medical Center or St. Luke's Hospital (St. Louis). The group representing Y-chromosomal diversity consisted largely of DNA samples purchased from the NHGRI/NIGMS DNA Polymorphism Discovery Resource[23] (Coriell Cell Repositories). **Supplementary Table 5** online lists the sources of the remaining samples in this group.

In the association study, both affected and unaffected men were from couples presenting with infertility at the Academic Medical Center (Amsterdam). A man was considered to be affected if he had a sperm count of $<10^7$ per ml or a total sperm count $<2 \times 10^7$. A man was considered to have normal spermatogenesis if his total sperm count was $>4 \times 10^7$ and sperm motility and morphology were normal by World Health Organization criteria[26].

In the initial screen and in the association study, we excluded men known to have had any of the following conditions or treatments: Y-chromosomal deletions known to be associated with spermatogenic failure (*AZFa*, *AZFc* or P5/P1 deletions; refs. 1,7,21); a 47,XXY karyotype; orchitis; cryptorchidism; radiotherapy; or chemotherapy. We prepared DNA from peripheral blood leukocytes or EBV-transformed lymphoblastoid cell lines.

Deletion screening. Using human genomic DNA as templates for PCR, we identified gr/gr deletions by the following STS results: sY1291 negative; and sY1161, sY1191, sY1206 and sY1201 all positive (**Fig. 1a**,**c**). We identified the b1/b3 deletion by the following STS results: sY1161, sY1197, sY1191 and sY1291 all negative; and sY142, sY1258, sY1206 and sY1201 all positive (**Fig. 2a**,**b**). See GenBank for PCR primers and conditions. We carried out two-color FISH analysis as described previously[27].

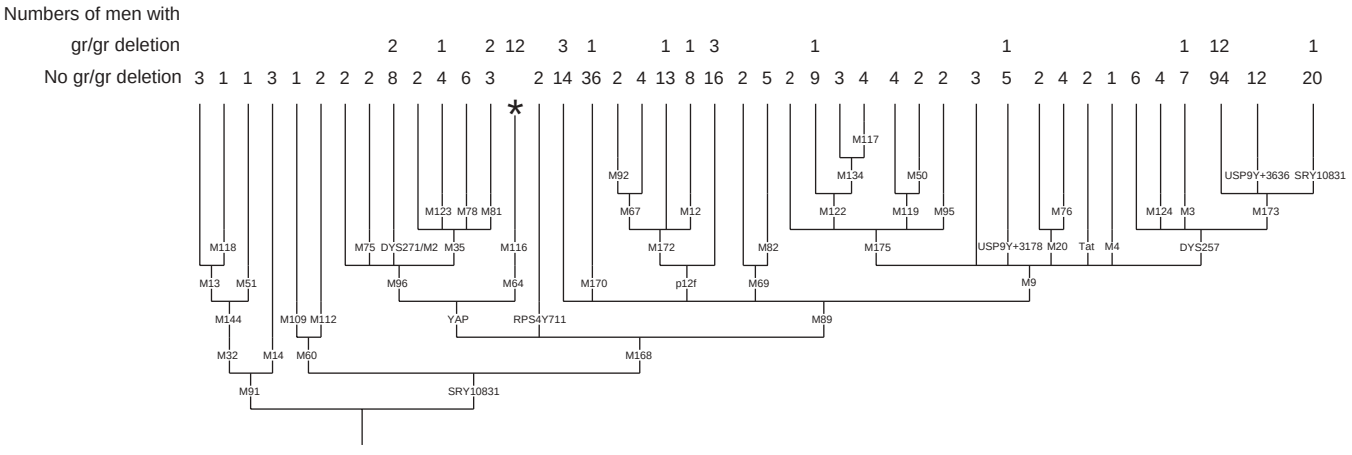

Figure 4 The genealogical tree of human Y chromosomes and branches in which the gr/gr deletion was observed. Individuals were assigned to one of 43 branches by typing for the stable, biallelic polymorphisms indicated (*e.g.*, M91, M60; see **Supplementary Table 6** online). Above each branch is shown the numbers of men studied and found to be gr/gr-deleted or to have no gr/gr deletion. Asterisk indicates branch D2b. All gr/gr-deleted Y chromosomes in this branch may be descendants of a single gr/gr-deleted founder (**Supplementary Note** online). In other branches, the proportions of gr/gr-deleted chromosomes reported here are probably higher than in the general population because many of the men studied had spermatogenic failure. **Supplementary Figure 1** online provides Y Chromosome Consortium[15] designations for branches with gr/gr deletions.

Y-chromosome haplotyping. Individuals were haplotyped using the Y-linked polymorphisms listed in **Supplementary Table 6** and **Supplementary Table 7** online.

GenBank accession numbers for STSs. sY142, G38345; sY1161, G66148; sY1191, G73809; sY1197, G67168; sY1201, G67170; sY1206, G67171; sY1258, G75499; sY1291, G72340.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank D. Altshuler, A. Chakravarti, A.G. Clark, G.Q. Daley, M.J. Daly, J.N. Hirschhorn, L. Kruglyak, V.K. Mootha, S. Paabo and D.E. Reich for comments on the manuscript; M.F. Hammer and P.A. Underhill for assistance with genealogical studies; T. Ogata and the Department of Endocrinology and Metabolism, National Research Institute for Child Health and Development, Japan, for support and advice; and C. Bruning, N.A. Ellis, S. Fallet, A. Garguilo, J. Gianotten, B.R. Gilbert, M.F. Hammer, W.A. Hogge, J. Hoo, T. Jenkins, K. Kepler, K. Monaghan, E. Pergament, B. Shapiro, M.C. Summers, U. Surti, L. Weiss and J. Weissenbach for DNA, cell and blood samples. This work was supported by the US National Institutes of Health, the Howard Hughes Medical Institute and the Academic Medical Center.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 18 April; accepted 16 September 2003
Published online at http://www.nature.com/naturegenetics/

1. Kuroda-Kawaguchi, T. *et al.* The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.* **29**, 279–286 (2001).
2. Yen, P. The fragility of fertility. *Nat. Genet.* **29**, 243–244 (2001).
3. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
4. Sun, C. *et al.* Deletion of azoospermia factor a (*AZFa*) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum. Mol. Genet.* **9**, 2291–2296 (2000).
5. Blanco, P. *et al.* Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.* **37**, 752–758 (2000).
6. Kamp, C., Hirschmann, P., Voss, H., Huellen, K. & Vogt, P.H. Two long homologous retroviral sequence blocks in proximal Yq11 cause *AZFa* microdeletions as a result of intrachromosomal recombination events. *Hum. Mol. Genet.* **9**, 2563–2572 (2000).
7. Repping, S. *et al.* Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am. J. Hum. Genet.* **71**, 906–922 (2002).
8. Jobling, M.A. *et al.* Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Hum. Mol. Genet.* **5**, 1767–1775 (1996).
9. Bienvenu, T., Patrat, C., McElreavey, K., de Almeida, M. & Jouannet, P. Reduction in the *DAZ* gene copy number in two infertile men with impaired spermatogenesis. *Ann. Genet.* **44**, 125–128 (2001).
10. de Vries, J.W.A. *et al.* Reduced copy number of *DAZ* genes in subfertile and infertile men. *Fertil. Steril.* **77**, 68–75 (2002).
11. Fernandes, S. *et al.* High frequency of *DAZ1*/*DAZ2* gene deletions in patients with severe oligozoospermia. *Mol. Hum. Reprod.* **8**, 286–298 (2002).
12. Ewis, A.A., Lee, J., Shinka, T. & Nakahori, Y. Microdeletions of a Y-specific marker, Yfm1, and implications for a role in spermatogenesis. *J. Hum. Genet.* **47**, 257–261 (2002).
13. Underhill, P.A. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
14. Underhill, P.A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**, 43–62 (2001).
15. The Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
16. Kuroki, Y. *et al.* Spermatogenic ability is different among males in different Y chromosome lineage. *J. Hum. Genet.* **44**, 289–292 (1999).
17. Hammer, M. *et al.* Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203 (2001).
18. Ke, Y. *et al.* African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* **292**, 1151–1153 (2001).
19. Jobling, M. & Tyler-Smith, C. The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.* **4**, 598–612 (2003).
20. Reijo, R. *et al.* Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat. Genet.* **10**, 383–393 (1995).
21. Vogt, P.H. *et al.* Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet.* **5**, 933–943 (1996).
22. Oates, R.D., Silber, S., Brown, L.G. & Page, D.C. Clinical characterization of 42 oligospermic or azoospermic men with microdeletion of the *AZFc* region of the Y chromosome, and of 18 children conceived via ICSI. *Hum. Reprod.* **17**, 2813–2824 (2002).
23. Collins, F.S., Brooks, L.D. & Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231 (1998).
24. Haldane, J.B.S. The rate of spontaneous mutation of a human gene. *J. Genet.* **31**, 317–326 (1935).
25. Soucie, J.M., Evatt, B., Jackson, D. & The Hemophilia Surveillance System Project Investigators. Occurrence of hemophilia in the United States. *Am. J. Hematol.* **59**, 288–294 (1998).
26. World Health Organization. WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction (Cambridge University Press, Cambridge, New York, 1992).
27. Saxena, R. *et al.* Four *DAZ* genes in two clusters found in the *AZFc* region of the human Y chromosome. *Genomics* **67**, 256–267 (2000).
28. Tilford, C. *et al.* A physical map of the human Y chromosome. *Nature* **409**, 943–945 (2001).