nature genetics

High mutation rates have driven extensive structural polymorphism among human Y chromosomes

Sjoerd Repping[1,2], Saskia K M van Daalen[2], Laura G Brown[1], Cindy M Korver[2], Julian Lange[1], Janet D Marszalek[1], Tatyana Pyntikova[1], Fulco van der Veen[2], Helen Skaletsky[1], David C Page[1] & Steve Rozen[1]

Although much structural polymorphism in the human genome has been catalogued[1–5], the kinetics of underlying change remain largely unexplored. Because human Y chromosomes are clonally inherited, it has been possible to capture their detailed relationships in a robust, worldwide genealogical tree[6,7]. Examination of structural variation across this tree opens avenues for investigating rates of underlying mutations. We selected one Y chromosome from each of 47 branches of this tree and searched for large-scale variation. Four chromosomal regions showed extensive variation resulting from numerous large-scale mutations. Within the tree encompassed by the studied chromosomes, the distal-Yq heterochromatin changed length ≥ 12 times, the *TSPY* gene array changed length ≥ 23 times, the 3.6-Mb IR3/IR3 region changed orientation ≥ 12 times and the *AZFc* region was rearranged ≥ 20 times. After determining the total time spanned by all branches of this tree (~ 1.3 million years or 52,000 generations), we converted these mutation counts to lower bounds on rates: $\geq 2.3 \times 10^{-4}$, $\geq 4.4 \times 10^{-4}$, $\geq 2.3 \times 10^{-4}$ and $\geq 3.8 \times 10^{-4}$ large-scale mutations per father-to-son Y transmission, respectively. Thus, high mutation rates have driven extensive structural polymorphism among human Y chromosomes. At the same time, we found limited variation in the copy number of Y-linked genes, which raises the possibility of selective constraints.

Recent studies point to substantial large-scale copy number variation within the human genome, and a few such studies have shown large inversions[1–5]. With the exception of one large inversion that arose once in human history[4], the mutational dynamics underlying common large-scale, structural polymorphisms have not been addressed. Are these polymorphisms usually the result of independent, recurrent mutation, or are they inherited from a single founder? How often do mutations generate structural variants? The male-specific region of the human Y chromosome (**Fig. 1**) offers unique opportunities for investigating these questions because of its clonal inheritance and the availability of a robust genealogical tree that describes in detail the relationships among extant Y chromosomes (**Fig. 2**)[6,7].

With these questions in mind, we assembled a collection of Y chromosomes representing 47 major branches of the genealogy and encompassing worldwide diversity (**Fig. 2** and **Supplementary Table 1** and **Supplementary Fig. 1** online). On the basis of analysis of published empirical evidence, we searched for nine broad categories of potential structural variation among these chromosomes (**Fig. 1a**, **2**, **3** and **Supplementary Methods** online). For several reasons, we focused on potential structural variation involving the segmentally duplicated, ampliconic regions of the Y chromosome. Previous results

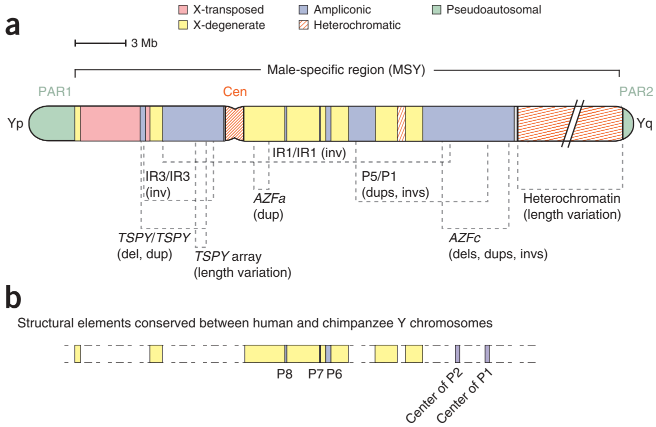

Figure 1 Overview of potential structural variation in the human Y chromosome. At top, the structure of the reference Y chromosome, including short and long arms (Yp and Yq), pseudoautosomal regions 1 and 2 (PAR1 and PAR2) and centromere (Cen). (**a**) Potential structural polymorphisms for which we assayed (details in **Supplementary Methods**). (**b**) Structural elements conserved between human and chimpanzee Y chromosomes are shown according to their position in the reference human Y chromosome. These conserved elements consist of the X-degenerate sequence, palindromes P8, P7 and P6 and the centers of palindromes P2 and P1.

[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 1105 AZ, The Netherlands. Correspondence should be addressed to D.C.P. (page_admin@wi.mit.edu).

Received 21 December 2005; accepted 31 January 2006; published online 26 February 2006; doi:10.1038/ng1754

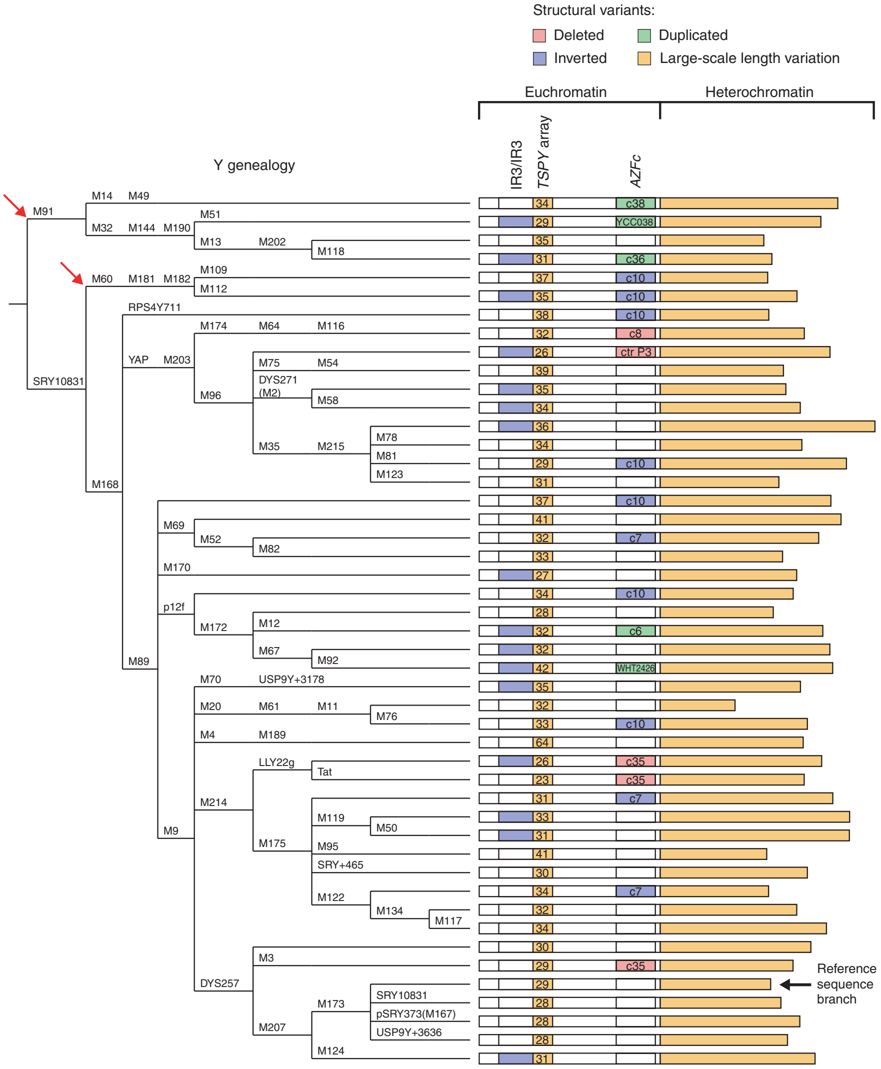

Figure 2 Y chromosome genealogical tree (left) and identified structural polymorphisms (right). Chromosomes were assigned to one of 47 branches by typing for the stable, biallelic polymorphisms indicated (for example, M91 and M60; refs. 6,7). Red arrows indicate major branches confined to Africa[6]. For each branch, the structure of the Y chromosome sampled is schematized, including, at far right, the length of distal-Yq heterochromatin. Within the euchromatin, the presence of a particular structural variant is indicated by a color-coded rectangle. Codes denoting specific *AZFc* architectures are explained in **Figure 4**, **Supplementary Table 2** and **Supplementary Figures 2–7**. See **Supplementary Figures 3** and **4** for the 'ctr P3' deletion and for 'YCC038', which contains a small deletion, but in which duplication predominates. The reference Y chromosome belongs to the indicated branch (**Supplementary Methods**), but, as no corresponding cell line exists, its heterochromatin and *TSPY* array lengths could not be determined. **Supplementary Figure 1** provides sample identifiers and Y-haplotype designations[6,7].

showed only a handful of large-scale structural rearrangements in nonampliconic portions of the human Y in the ∼6.5 million years since humans and chimpanzees diverged[8,9]. Moreover, none of these differences is polymorphic among extant human Y chromosomes (data reported in **Supplementary Methods**). At the same time, available data suggest that there is little large-scale structural similarity between the ampliconic regions of the human and chimpanzee Y chromosomes, with conserved ampliconic structures confined to palindromes P6, P7 and P8 and the centers of palindromes P1 and P2 (refs. 9,10; **Fig. 1b**). Thus, among human Y chromosomes, structural polymorphisms would most likely involve ampliconic regions.

We designed assays that would be maximally informative for each of the nine categories of structural variation, which include inversions and subtle changes in copy number (**Supplementary Methods**). For example, pulsed-field DNA blots can detect subtle differences in the length of the *TSPY* array[11], metaphase FISH can detect several different kinds of potential pericentric inversions[12,13] and multicolor interphase FISH can detect other large inversions and distinguish between alternative large-scale organizations of the *AZFc* region[14,15].

Of the nine categories of potential structural variation, our search detected four, which occurred in four regions of the chromosome. Two of the regions, the distal-Yq heterochromatin and the *TSPY* array, showed large-scale length variation. The distal-Yq heterochromatin is composed of low-complexity sequences organized in tandem arrays[16]. It ranged in length from 29% to 54% of the metaphase Y chromosome, with a median of 44% (**Figs. 1**, **2**, **3a** and **5a**). The *TSPY* array is composed of highly similar 20.4-kb repeat units, each containing a copy of the *TSPY* gene and of the *CYorf16* transcription unit[11,16,17]. The *TSPY* array ranged in size from 23 to 64 units (0.47 to 1.3 Mb), with a median of 32 units (0.65 Mb; **Figs. 1**, **2**, **3b** and **5b**).

The third region, in proximal Yp, was inverted in 16 chromosomes (**Figs. 1**, **2**, **3c–f**, **5d**)[18,19]. We localized the boundaries of this 3.6-Mb inversion to within 100 kb of the IR3 repeats, strongly supporting the model that the inversions originated via ectopic homologous recombination between the IR3 repeats (**Fig. 3c–f**)[16,20].

The fourth region, *AZFc*, demonstrated abundant architectural polymorphism (**Figs. 1**, **2**, **4** and **5d**). Because this region is composed almost entirely of large, nearly identical, repeated amplicons[21], there are myriad possibilities for rearrangement via ectopic homologous recombination. We predicted *AZFc* architectures that could result from homologous recombination between amplicons ≥100 kb in length and designed combinations of two-color FISH and plus/minus STSs to detect these potential architectures (**Supplementary Table 2** and **Supplementary Fig. 2** online). These assays showed that 20 of the 47 chromosomes had variant *AZFc* architectures, the largest of which involved a duplication of ∼3.5 Mb (**Figs. 2**, **4**, **5d** and **Supplementary Figs. 3–7** online).

For each of the four regions showing structural polymorphism, we determined the minimum number of independent mutation events needed to produce the distribution of variants across the genealogical tree; that is, a minimum-mutation history. The many distinct lengths that we observed in the distal-Yq heterochromatin and *TSPY* array must have been the result of multiple mutations (**Figs. 2** and **5a**,**b**). For a more complete analysis, we calculated minimum-mutation histories using methods that accommodate experimental variance in

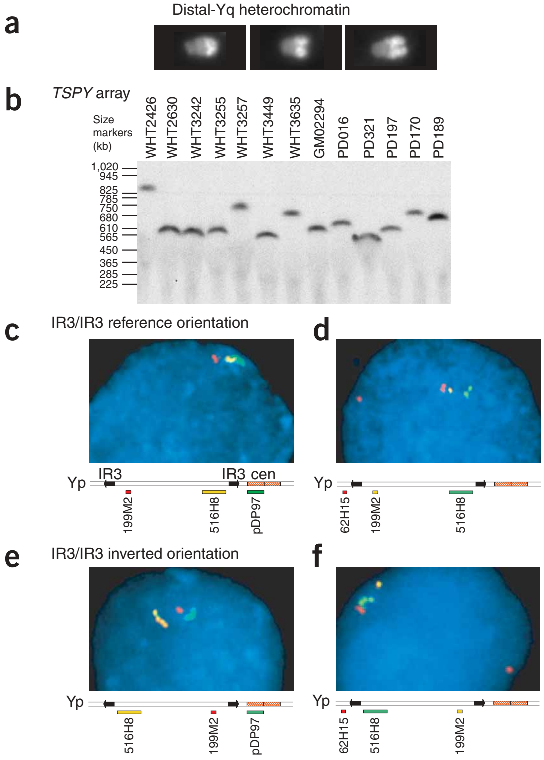

Figure 3 Assaying variation in heterochromatin length, *TSPY* array length and IR3/IR3 orientation. (**a**) Quinacrine staining of metaphase Y chromosomes with distal-Yq heterochromatin that is short (sample 4566), average (PD178) or long (PD123). (**b**) *Pme*I pulsed-field DNA blot to assay the number of *TSPY* repeats. (**c**,**d**) Three-color FISH of interphase nuclei with the reference orientation of the IR3/IR3 region (sample WHT3242). Below each nucleus is a schematic diagram of proximal Yp with IR3 repeats indicated; regions detected by each probe (199M2, 516H8, pDP97 and 62H15) are indicated in the color of the probe's stain. (**e**,**f**) Interphase nuclei with IR3/IR3 inversion (sample WHT3257). Results shown in **c** and **e** map the proximal inversion breakpoint between 516H8 and pDP97 in the reference orientation and between 199M2 and pDP97 in the inverted orientation. Results shown in **d** and **f** map the distal inversion breakpoint between 62H15 and 199M2 in the reference orientation and between 62H15 and 516H8 in the inverted orientation. 62H15 cross-hybridizes to the X chromosome, generating additional red dots at nuclear margins in **d** and **f**.

allele length (**Supplementary Methods**)[22]. For the *TSPY* array, this yielded an estimate of ≥23 changes in length, and for the distal-Yq heterochromatin, an estimate of ≥12 large changes in length. For the *TSPY* array, these changes were probably due to unequal crossing-over. Likewise, for the distal-Yq heterochromatin, large changes were probably attributable to unequal crossing-over, although small changes probably also occurred via mutational mechanisms that operate on micro- or minisatellites. For IR3/IR3, there were ≥12 independent inversion events, and for *AZFc*, there were ≥20 rearrangement events (**Supplementary Methods**). We also noted that in minimum-mutation histories of *AZFc*, inversion events were overrepresented compared with a null model of equally probable inversions, deletions and duplications ($P < 0.038$; **Supplementary Methods**). A predominance of inversion events could be caused by (i) more frequent inversion events than deletion or duplication events or (ii) natural selection against deletions and duplications but not inversions.

Having estimated lower bounds on the numbers of mutations causing structural variation, we proceeded to investigate their rates. For this, we needed to determine the denominator of the rate: that is, the total time represented by all branches in the genealogy. To estimate this, we used the total number of SNPs in the tree, the average number of SNPs on paths from the root to the leaves and the time to the last common ancestor of extant human Y chromosomes (**Supplementary Methods**). Use of previously reported SNPs in this estimate might have led to bias, if some parts of the tree were more intensively screened for SNPs than others. Therefore, we resequenced ~80 kb in the 47 chromosomes, thereby identifying 94 SNPs in an unbiased way. Using these SNPs, we estimated the total time represented by the

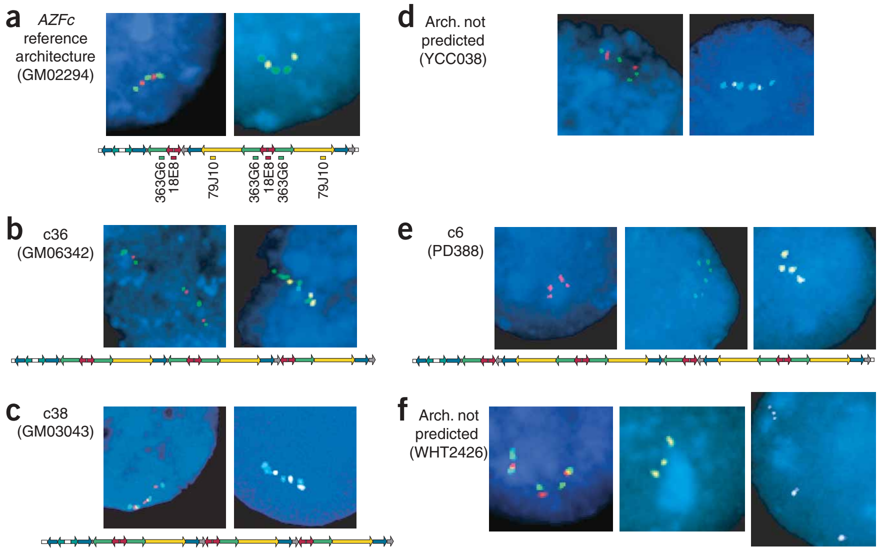
a
AZFc reference architecture (GM02294)
363G6
18E8
79J10
363G6
18E8
363G6
79J10
b
c36 (GM06342)
c
c38 (GM03043)
d
Arch. not predicted (YCC038)
e
c6 (PD388)
f
Arch. not predicted (WHT2426)

Figure 4 Detecting architectural variation in *AZFc*. Sample identifiers shown in parentheses. (**a**) Two-color FISH of interphase nuclei with *AZFc* reference architecture. Below the nuclei, *AZFc* reference architecture is depicted as a sequence of color-coded arrows representing amplicons[21]. Probes and sites of hybridization are shown; probe colors match those of detected amplicons. Left: FISH with green and red probes. Right: FISH with green and yellow probes. (**b**–**d**) Interphase nuclei with variant *AZFc* architectures probed as in **a**. Inferred amplicon organizations are shown below pairs of nuclei. (**e**) Interphase nuclei probed with 18E8 (red, left), 363G6 (green, center) and 79J10 (yellow, right), indicating four pairs of red amplicons, six green amplicons and four yellow amplicons. Of *AZFc* architectures with these counts, only c6 is likely to be generated from the reference by one recombination event, although others can be generated by two successive events (**Supplementary Figs. 2** and **5** and **Supplementary Table 2**). (**f**) Interphase nuclei from sample WHT2426. Left: probed as in left panel of **a**; center: probe 79J10; right: probe 366C6, which hybridizes to the gray amplicon in *AZFc* (see **a**) and to chromosome 1 (**Supplementary Methods**). Three closely spaced dots at the upper left arise from *AZFc* and indicate three gray amplicon copies. The two strong signals in the lower half arise from chromosome 1. No predicted *AZFc* architecture would yield this pattern of FISH results (**Supplementary Table 2**).

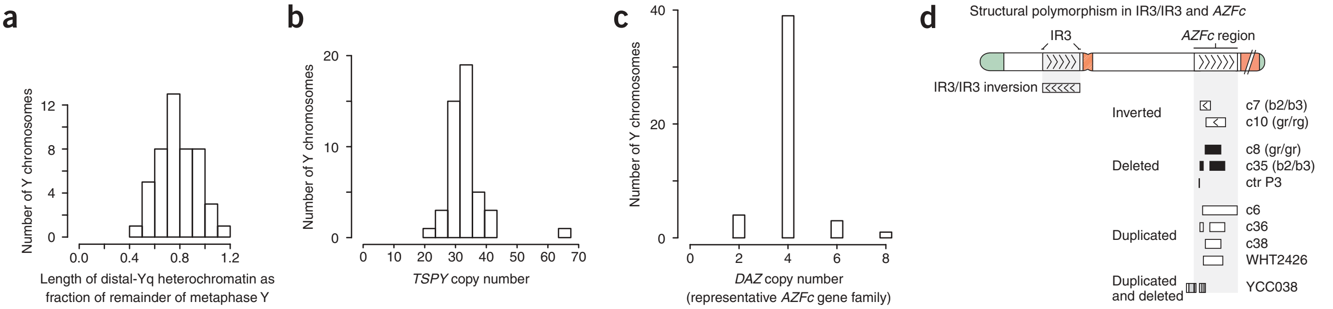

Figure 5 Summary of identified Y chromosome structural variation. (**a**–**c**) Distributions of (**a**) heterochromatin length, (**b**) number of repeat units in the *TSPY* array[16] and (**c**) *DAZ* gene copy number. See **Supplementary Figure 8** for *CDY1* and *BPY2* gene copy numbers. (**d**) IR3/IR3 inversion in Yp and variant architectures involving *AZFc*. The sizes and locations of the duplications and deletions in YCC038 and WHT2426 are estimated (**Fig. 4d**,**f**, **Supplementary Table 1**, **Supplementary Fig. 4** and **Supplementary Methods**).

genealogical tree at 1.3 million years, which conservatively corresponds to 52,000 generations (**Supplementary Methods**).

Using this denominator and mutation counts from the minimum-mutation histories, we inferred lower bounds on mutation rates. These are lower bounds because such histories never involve reversion or recurrence events unless essential to explain the distribution of variants. For the distal-Yq heterochromatin, ≥12 large changes in length over 52,000 generations correspond to a rate $\geq 2.3 \times 10^{-4}$ large-scale mutations per father-to-son transmission of a Y chromosome. For the *TSPY* array, ≥23 changes in length correspond to a rate $\geq 4.4 \times 10^{-4}$. For *AZFc*, ≥20 rearrangement events in the tree correspond to a rate $\geq 3.8 \times 10^{-4}$, a lower bound broadly consistent with the independent estimate of 2.5×10^{-4} for one particular *AZFc* mutation, the b2/b4 deletion[21]. For IR3/IR3, the minimum-mutation count of ≥12 inversion events corresponds to a rate of $\geq 2.3 \times 10^{-4}$. In addition, it was possible to obtain a maximum-likelihood estimate of the rate of IR3/IR3 inversion events (**Supplementary Methods**). These events seem to have resulted from a single mutational mechanism and are likely to have occurred at the same rate regardless of the orientation of the IR3/IR3 region. Thus, the analysis needed to examine only a single parameter, the rate of inversions, which showed maximum likelihood at 9.2×10^{-4} inversion events per father-to-son transmission of a Y chromosome.

How do the rates of large-scale structural mutation estimated here compare with rates of other kinds of mutations in the human genome? The rates we observed are at the low end of the range of rates among mini- and microsatellites but are ~10,000 times the average rate of single-nucleotide substitutions (**Supplementary Methods**). Considering that structural mutations of the Y chromosome often affect hundreds or thousands of kilobases and sometimes alter gene copy number, these mutations may be a major source of Y-linked phenotypic variation in human populations.

Despite the prevalence of Y chromosomes with large-scale differences from the reference sequence (**Fig. 2**), multicopy gene families showed limited variation in copy number, with pronounced modes and few excursions to low or high numbers of copies. We observed gene copy number variation only in the *TSPY* array and in *AZFc* and flanking areas. In *AZFc*, a predominance of inversions resulted in few chromosomes with gene copy numbers that differed from the reference sequence (**Figs. 2**, **5c** and **Supplementary Fig. 8**). Furthermore, the *TSPY* genes, whose tandem array has undergone frequent changes in length, also showed limited variation in copy number. Indeed, the coefficient of variation of *TSPY* copy number (18.6%; s.d. as a percent of the mean) was less than that of *AZFc* gene families (*DAZ*, 24.2%; *BPY*, 27.4%; *CDY1*, 24.2%; **Fig. 5b**,**c** and **Supplementary Fig. 8**). Is limited variation in gene copy number consistent with the high mutation rates underlying widespread structural diversity among human Y chromosomes? As previously reported, natural selection has acted against one *AZFc* variant, the gr/gr deletion, in which several testis-specific gene families have reduced copy numbers and which confers increased risk of spermatogenic failure[14,23–27]. Thus, it is possible that natural selection had a wider role in constraining variation among human Y chromosomes by removing extremely high– or low–copy number variants from the population.

METHODS

Human samples. All assays were performed on human lymphoblastoid cell lines, cultured human fibroblasts or DNAs extracted from them. Most of these samples were obtained from the National Human Genome Research Institute/National Institute of General Medical Sciences DNA Polymorphism Discovery Resource (Coriell Cell Repositories)[28] or other public collections. To maximize coverage of the Y chromosome genealogical tree, we also studied several human cell lines from our own collections (**Supplementary Fig. 1** and **Supplementary Table 1**). **Supplementary Methods** lists availability of cell lines representing the structural variants described here.

Length of distal-Yq heterochromatin. For 46 of 47 men tested, we used quinacrine staining to measure heterochromatin length as a fraction of the total length of the metaphase Y chromosome, as previously described[29] (≥25 nuclei per sample, except for WHT3870 (12 nuclei)). We assessed the reproducibility of these measurements as discussed in **Supplementary Methods**. In one sample, WHT3299, the Y chromosome contained so little distal-Yq heterochromatin that it could not be measured using quinacrine staining. Instead, we used metaphase FISH (**Supplementary Methods**). The very short heterochromatin in individual WHT3299 was inherited by his son and thus was not an artifact of cell culture.

Length of *TSPY* array. We used *Pme*I pulsed-field DNA blotting to measure the length of the *TSPY* array (**Fig. 3b**). Gels were electrophoresed for 25 h at 14 °C, 6 V cm^{-1} (200 V), with a 60- to 120-s switch-time ramp. The probe was the PCR product of STS sY1256. We estimated the number of *TSPY* repeats by subtracting the lengths of non–*TSPY*-repeat flanks (10.9 kb) at the ends of the *Pme*I fragment, dividing by 20.37 kb (the size of the repeat unit[16]) and rounding. In all chromosomes, we confirmed, by sequencing, the presence of the *Pme*I site proximal to the *TSPY* array (**Supplementary Methods**). We did not sequence the *Pme*I site distal to the array, but loss of that site would increase the size of the *Pme*I fragment by only 16 kb. To further verify our findings, we assayed all samples on DNA blots based on a second restriction enzyme, *Xba*I (ref. 11) and obtained size estimates consistent with the *Pme*I-based sizes (**Supplementary Methods**).

Detecting IR3/IR3 orientation and *AZFc* architectures. One-, two- or three-color FISH was performed as described[30]. For each sample and set of probes,

≥200 nuclei were scored. Apart from plasmid pDP97 and cosmid 18E8, all clones used as probes were BACs derived from the RPCI-11 library (prefix 'RP11-'); refs. 16,20 provide map positions. The computer program that enumerated potential *AZFc* architectures is available on request (S.R.).

Rates of mutations giving rise to structural polymorphism. We determined minimum-mutation histories of the structural variants studied either manually (IR3/IR3 orientation, *AZFc* architecture) or using our implementation of Sankoff's algorithm[22] (distal-Yq heterochromatin, *TSPY* array, IR3/IR3 orientation); code is available on request (S.R.). We sequenced 237 PCR products in the 47 chromosomes to ascertain in an unbiased way the SNPs used to determine the total length of time represented by the tree. All SNPs detected, as well as the genotypes of the 47 chromosomes at these SNPs, have been submitted to dbSNP. See **Supplementary Methods** for details of the maximum likelihood analysis of the rate of IR3/IR3 inversion events.

Accession codes. GenBank: PCR product of STS sY1256, G75613; PCR products resequenced in SNP discovery, BV678971–BV679207.

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

ACKNOWLEDGMENTS

We thank G. Farino for DNA sequencing; V. Frazzoni and G. Rogers for tissue culture; N.A. Ellis, M.F. Hammer, T. Jenkins, R.D. Oates and S. Silber for cell lines and blood samples; J. de Vries, N. Leschot and P. Underhill for technical and scientific advice; J.E. Alfoldi, A.E. Baltus, D.W. Bellott, A. Chakravarti, M.J. Daly, J.F. Hughes, L. Kruglyak, Y.-H. Lim, J.L. Mueller and D.E. Reich for comments on the manuscript and A.G. Clark for advice and guidance on studies of mutation rates. This work was supported by the US National Institutes of Health, the Howard Hughes Medical Institute, the Netherlands Organization for Scientific Research and the Academic Medical Center.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
2. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
3. Sharp, A.J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
4. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).
5. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **37**, 727–732 (2005).
6. Underhill, P.A. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
7. The Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
8. Vignaud, P. *et al.* Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature* **418**, 152–155 (2002).
9. Hughes, J.F. *et al.* Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437**, 100–103 (2005).
10. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
11. Tyler-Smith, C., Taylor, L. & Muller, U. Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *J. Mol. Biol.* **203**, 837–848 (1988).
12. Grace, H.J., Ally, F.E. & Paruk, M.A. 46,Xinv(Yp+q-) in four generations of an Indian family. *J. Med. Genet.* **9**, 293–297 (1972).
13. Bernstein, R., Wadee, A., Rosendorff, J., Wessels, A. & Jenkins, T. Inverted Y chromosome polymorphism in the Gujerati Muslim Indian population of South Africa. *Hum. Genet.* **74**, 223–229 (1986).
14. Repping, S. *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**, 247–251 (2003).
15. Repping, S. *et al.* A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. *Genomics* **83**, 1046–1052 (2004).
16. 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).
17. Manz, E., Schnieders, F., Muller Brechlin, A. & Schmidtke, J. *TSPY*-related sequences represent a microheterogeneous gene family organized as constitutive elements in *DYZ5* tandem repeat units on the human Y chromosome. *Genomics* **17**, 726–731 (1993).
18. Affara, N.A. *et al.* Variable transfer of Y-specific sequences in XX males. *Nucleic Acids Res.* **14**, 5375–5387 (1986).
19. Page, D.C. Sex reversal: deletion mapping the male-determining function of the human Y chromosome. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 229–235 (1986).
20. Tilford, C. *et al.* A physical map of the human Y chromosome. *Nature* **409**, 943–945 (2001).
21. 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).
22. Sankoff, D. Minimal mutation trees of sequences. *SIAM J. Appl. Math.* **28**, 35–42 (1975).
23. Machev, N. *et al.* Sequence family variant loss from the AZFc interval of the human Y chromosome, but not gene copy loss, is strongly associated with male infertility. *J. Med. Genet.* **41**, 814–825 (2004).
24. de Llanos, M., Ballesca, J.L., Gazquez, C., Margarit, E. & Oliva, R. High frequency of gr/gr chromosome Y deletions in consecutive oligospermic ICSI candidates. *Hum. Reprod.* **20**, 216–220 (2005).
25. Ferlin, A. *et al.* Association of partial *AZFc* region deletions with spermatogenic impairment and male infertility. *J. Med. Genet.* **42**, 209–213 (2005).
26. Hucklenbroich, K. *et al.* Partial deletions in the *AZFc* region of the Y chromosome occur in men with impaired as well as normal spermatogenesis. *Hum. Reprod.* **20**, 191–197 (2005).
27. Lynch, M. *et al.* The Y chromosome gr/gr subdeletion is associated with male infertility. *Mol. Hum. Reprod.* **11**, 507–512 (2005).
28. 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).
29. Schnedl, W. Flurescenzuntersuchungen ueber die langenvariabilitaet des Y-chromosoms beim menschen. *Humangenetik* **12**, 188–194 (1971).
30. 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).