

Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing *Amelogenin Y*

Mark A. Jobling^{*1}, Iek Chi C. Lo¹, Daniel J. Turner², Georgina R. Bowden¹, Andrew C. Lee¹, Yali Xue², Denise Carvalho-Silva², Matthew E. Hurles², Susan M. Adams¹, Yuet Meng Chang³, Thirsa Kraaijenbrink⁴, Jürgen Henke⁵, Ginevra Guanti⁶, Brian McKeown⁷, Roland A.H. van Oorschot⁸, R. John Mitchell⁹, Peter de Knijff⁴, Chris Tyler-Smith² and Emma J. Parkin¹

¹ Department of Genetics, University of Leicester, University Road, Leicester, UK

² Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK

³ Malaysian Forensic DNA Laboratory, Department of Chemistry, Jalan Sultan, Selangor, Malaysia

⁴ MGC-Department of Human and Clinical Genetics, Leiden University Medical Centre, The Netherlands

⁵ Institut für Blutgruppenforschung LGC GmbH, Köln, Germany

⁶ Cattedra di Genetica Medica, Policlinico, Piazza Giulio Cesare, Bari, Italy

⁷ Orchid Cellmark Ltd, Blacklands Way, Abingdon, UK

⁸ Victoria Police Forensic Services Department, Melbourne, Victoria, Australia

⁹ Department of Genetics and Human Variation, School of Molecular Sciences, La Trobe University, Melbourne, Australia

*Corresponding author:

Prof Mark A. Jobling, Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Tel.: +44 (0)116 252 3427 Fax: +44 (0)116 252 3378

Email: maj4@le.ac.uk

© 2006 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Structural polymorphism is increasingly recognised as a major form of human genome variation, and is particularly prevalent on the Y chromosome. Assay of the *Amelogenin Y* gene (*AMELY*) on Yp is widely used in DNA-based sex testing, and sometimes reveals males who have interstitial deletions. In a collection of 45 deletion males from 12 populations, we used a combination of STS (sequence-tagged site) mapping, and binary-marker and Y-STR (short tandem repeat) haplotyping to understand the structural basis of this variation. 41/45 males carry indistinguishable deletions, 3.0-3.8Mb in size. Breakpoint mapping strongly implicates a mechanism of non-allelic homologous recombination between the proximal major array of *TSPY*-gene-containing repeats, and a single distal copy of *TSPY*; this is supported by estimation of *TSPY* copy number in deleted and non-deleted males. The remaining four males carry three distinct non-recurrent deletions (2.5-4.0Mb) which may be due to non-homologous mechanisms. Haplotyping shows that *TSPY*-mediated deletions have arisen seven times independently in the sample. One instance, represented by 30 chromosomes mostly of Indian origin within haplogroup J2e1*/M241, has a time-to-most-recent-common-ancestor of $\sim 7700 \pm 1300$ years. In addition to *AMELY*, deletion males all lack the genes *PRKY* and *TBL1Y*, and the rarer deletion classes also lack *PCDH11Y*. The persistence and expansion of deletion lineages, together with direct phenotypic evidence, suggests that absence of these genes has no major deleterious effects.

Introduction

Recent studies have revealed a previously unexpected degree of segmental duplication (1), structural polymorphism and copy number variation (reviewed in ref. (2) in the human genome. Among human chromosomes, the Y bears a particularly large proportion of segmental duplications (3), and has long been known to exhibit a high degree of cytogenetically visible structural polymorphism, including heterochromatin and euchromatin length variation (4, 5), inversions (6, 7), and neutral translocations with autosomes (8). Unlike other human chromosomes, the Y has no requirement for pairing along its entire length with a homologue, and this may explain its structurally variable nature. A recent phylogenetically based molecular survey has emphasised the dynamic nature of the chromosome (9), but, other than the well-established variation in copy number at the *TSPY* tandem array (10) and 3.6-Mb inversion polymorphism (11, 12), discovered no structural variation on Yp.

One locus on Yp that is, in effect, routinely assayed on a large scale for presence-absence polymorphism is the *Amelogenin Y* gene (*AMELY*), since it forms part of the DNA-based sex test routinely employed in prenatal diagnosis (13), forensic typing (14), archaeological analysis (15, 16), and paternity testing. The test relies on the simultaneous PCR amplification of differently sized X- and Y-specific fragments from the XY-homologous amelogenin gene pair (14), lying in Xp22.3 and Yp11.1; however, some males have been reported in whom *AMELY* fails to amplify, and who, on DNA evidence alone, might therefore be misinterpreted as females. Initially, two such males were identified in a sample of 24 Sri Lankans (17), who lacked *AMELY* through a deletion on the short arm of the Y chromosome. The widespread use of the amelogenin PCR assay provides strong ascertainment of males with apparent interstitial *AMELY* deletions, and 65 of these have subsequently been reported (18-30). Although limited haplotyping and physical mapping has been undertaken in some

of these studies, no systematic analysis of the origin and extent of these structural variants in a substantial collection of chromosomes has been carried out.

The availability of the near-complete sequence of a Y chromosome (3) allows the investigation of the mutational mechanism(s) underlying deletions: in the original cases described (17), evidence from Southern blotting analysis of pulsed-field gels was used to propose a mechanism of non-allelic homologous recombination (NAHR) between copies of the 20.4-kb repeat units containing *TSPY* genes on Yp. We can now ask whether this mechanism is supported by breakpoint mapping, and whether it is responsible for recurrent deletions in an analogous way to the repeat-mediated deletions of the *AZFa* (31-33), *AZFb* (34) and *AZFc* (35) regions on Yq, associated with male infertility. Aside from their forensic relevance, *AMELY* deletion males are of phenotypic interest because they lack the *Amelogenin Y* gene itself, and may also lack other nearby genes on Yp. A study of such males could throw light on the selective influence that the absence of any such genes might have.

Here we describe a collection of 45 males from 12 different populations carrying *AMELY* deletions, and use a combination of sequence-tagged-site (STS) deletion mapping, binary-marker and Y-STR haplotyping, and *TSPY* copy number estimation to ask how many independent deletion events are represented in this sample. Based on the Y chromosome reference database sequence (3) we also estimate the sizes of the deletions and the number of genes deleted, and propose molecular mechanisms for their generation.

Results

A collection of 45 DNA samples from males showing amplification of only the X-specific amelogenin PCR product was assembled (*Figure 1*). Two of these males, m632 and m640, were the first deletion individuals to be described (17), and 30 others have also been previously reported (19, 22, 23, 25, 26, 28, 29). The remaining thirteen (see *Figure 1*) were discovered when the amelogenin sex test was applied either as part of an autosomal STR multiplex for paternity testing or DNA profiling, or as part of a Y-specific STR multiplex in population studies.

Deletion mapping

To estimate the sizes of the deletions, to ask what genes are co-deleted with *AMELY*, and to throw light on the underlying molecular mechanisms, we undertook deletion mapping using sequence-tagged sites (STSs) from Yp. FISH-based mapping approaches were precluded by the unavailability of cellular material. Many published STSs were initially developed to aid in clone contig construction (3, 36, 37), and therefore do not necessarily represent single-copy Y-specific sequences; they often detect homologs (containing identical primer binding sites) on the X chromosome, on autosomes, or elsewhere on the Y chromosome, making them useless for deletion mapping purposes in genomic DNA. STSs chosen around *AMELY* were therefore first verified as single-copy by BLAST searching, then tested for Y-specificity in normal male and female DNA samples before use. Within regions of high XY homology, novel STSs were designed near to existing ones, targeting Y-specific bases in XY sequence alignments; these are given the original STS names, with the suffix 'Y' (e.g. sY872Y).

Mapping using a total of 33 Y-specific markers distinguished five apparent different deletion classes (referred to here as Class I, I-sY59, II, III and IV) among the

45 chromosomes (*Figure 1*), suggesting at least five underlying events. The Class I deletion is by far the commonest, being found in 38 chromosomes (84%). It has its proximal breakpoint within the *TSPY* major array, as indicated by absence of the STS sY1079, lying just distal to the array, and the presence of sY59, lying at the proximal end. Its distal breakpoint lies in a 298-kb interval corresponding to the distal IR3 inverted repeat element, as shown by absence of the proximal marker sY1242 and presence of the distal sY1241. Further refinement of the distal breakpoint is rendered impractical by the presence of the second, highly homologous (99.75%; (3) copy of IR3, lying proximal to the deleted segment on Yp. The distal IR3 element contains a copy of the 20.4-kb *TSPY* repeat in the same orientation at the major *TSPY* array, with overall sequence similarity (considering the most distal copy in the array) of 96%, and a longest block of sequence identity of 796bp. The most parsimonious mechanism for this deletion is therefore NAHR between *TSPY* repeat copies, as originally proposed by Santos et al. (17). Based on the reference sequence, the extent of deleted DNA is between 3.0 and 3.8 Mb, depending on the position of the breakpoint in the major array. Individual deletion sizes are likely to vary further, due to differences in *TSPY* copy number.

The second apparent category of deletion ('Class I-sY59'), carried by three individuals, is related to the first. The distal breakpoint is the same as in Class I, but, proximally, sY59 is absent in addition to sY1079. Inspection of the sequence around sY59 shows that it lies just inside the proximal end of the *TSPY* array, between the penultimate and last copies. This might imply that the recombination event giving rise to these deletions occurred between the *TSPY* copy in the distal IR3 element and the most proximal copy of *TSPY* within the major array, to leave only a single copy of the repeat unit; however, when we screened a global collection of 188 *AMELY+* males with sY59 we found the marker to be absent in two cases, belonging to haplogroups I

and E(xE3b3). Presence/absence polymorphism of sY59 presumably arises through recombination within the proximal end of the *TSPY* array, and so its absence in some *AMELY*-deletion males is uninformative about the distal breakpoint. Deletions lacking sY59 should be classified simply as Class I, and therefore a total of 41/45 chromosomes belong to this class (91%).

The Class II deletion occurs in two individuals, and the remaining two classes (III and IV) are found as singletons (*Figure 1*). These three deletion classes are respectively 3.5-3.9Mb, 3.9-4.0Mb, and 2.5-3.1Mb in size, with uncertainty introduced by marker spacings at the breakpoints. To address the mechanism of deletion in these classes, we searched the breakpoint intervals in each class using the program Reputer (38) for perfect direct repeats that might be involved in sponsoring deletions. This revealed no repeats larger than the 'minimum efficient processing segment' length for homologous recombination of ~200bp (39), implying that the mechanism in all three classes may be non-homologous end-joining. However, there are many smaller perfect repeats forming parts of isolated SINEs, including one of 139bp in the Class IV breakpoint intervals, and these may have been involved in the deletion events; studies of some multi-megabase non-recurrent deletions on 17p11.2 have mapped breakpoints to a sequence identity block as small as 21bp within a SINE (40).

Haplotyping

Having identified four distinct deletion classes, we can ask if these reflect only four deletion events, or if some are recurrent. To do this we exploit the availability of a well resolved Y-chromosomal phylogeny based on binary markers (41, 42): if two deletion chromosomes within Class I or II belong to two different haplogroups within the phylogeny, the deletions are likely to have arisen independently. On the basis of

binary marker typing (*Figure 1, 2a*), Class I deletions are found in six different haplogroups showing that this *TSPY*-mediated deletion event is recurrent. Haplogroups D and G are each represented once, haplogroup I twice, haplogroup H(xH2) three times, and haplogroup R1b3 four times. The remaining 30 Class I chromosomes all belong to a single well-defined haplogroup, J2e1*/M241 (*Figure 1*). Class II deletions are present in two chromosomes, but both belong to hgR1b3, so these could represent identity by descent. The Class III deletion belongs to hgR1b3, and the Class IV deletion to hgJ(xJ2). The four deletion classes are therefore due to a minimum of nine events, and include a common founder *AMELY* Class I deletion lineage within haplogroup J2e1*/M241 - a finding consistent with previous studies that have identified deletions in this haplogroup (29, 30).

Y-STR analysis provides further haplotyping resolution, and clustering of haplotypes within a haplogroup can support the idea of common ancestry. We therefore determined 26-locus Y-STR haplotypes (26) for the deletion chromosomes (*Supplementary Table 2*). None of the 26 STRs typed fall within the deleted intervals of these chromosomes, though we note that DYS458, included in the widely used commercial forensic profiling kit Y-filer (Applied Biosystems), is absent in Class I deletions (*Figure 1; (29)*). *Figure 2b* shows a median joining network connecting the haplotypes, in which three clusters are evident. All four Class I deletion chromosomes belonging to hgR1b3 share a single haplotype, and the three Class I chromosomes within hgH(xH2) carry very closely related haplotypes, suggesting identity-by-descent and a single deletion event within each of these sets. The major cluster is formed by the hgJ2e1*/M241 chromosomes, and applying the rho statistic within Network yields a TMRCA (time-to-most-recent-common-ancestor) of $\sim 7700 \pm 1300$ years. The two deletion chromosomes within hgI have very different haplotypes, suggesting probable independence of origin; overall, then, Class I deletions appear to

have arisen on seven independent occasions in the set of 45 *AMELY*-deleted males. The two Class II deletion chromosomes both lie within hgR1b3, and the nature of the deletion they share, together with a shared Italian origin, strongly suggests common ancestry; their Y-STR haplotypes differ by only four mutational steps in the network, which is consistent with this. Considering all chromosomes, the 45 cases can be explained by a total of ten independent deletion events in four different classes.

***TSPY* copy number**

If the Class I deletions were indeed caused by *TSPY*-mediated NAHR, we might expect the copy number of *TSPY* repeats in deletion chromosomes to be reduced with respect to related non-deleted examples, the degree of reduction depending on the position of the breakpoint within the *TSPY* array. To investigate this, we carried out quantitative PCR to estimate *TSPY* copy number in the nine Nepalese Class I deletion chromosomes. For comparison, we also analysed the four available chromosomes from the same population and haplogroup (J2e1*/M241) that were not deleted for *AMELY*. Copy number estimation in other chromosomes was not possible due to scarcity or poor quality of genomic DNA.

Figure 3 shows *TSPY* copy number in deleted and non-deleted chromosomes within hgJ2e1*/M241, and includes two deletion cases, m632 and m640, for which *TSPY* copy number could be estimated from the size of the hybridising *Xba*I restriction fragment in pulsed-field gel analysis (17). The four non-deleted chromosomes have copy numbers ranging from 29-35 (mean 33), while the 11 deleted chromosomes have copy numbers ranging from 20-25 (mean 21). This difference is significant ($P = 0.003$, Mann-Whitney U test) and supports the *TSPY*-mediated recombination mechanism for deletions. Its simplest interpretation is that the founding *AMELY* deletion was accompanied by the loss of ~12 *TSPY* repeat units.

Notably, H401, the only Class I sY59-deleted chromosome in this analysis, has a copy number of 22, rather than the single copy expected if absence of both sY59 and *AMELY* were due to a single deletion event. This confirms that absence of sY59 is an independent event due to mutation within the *TSPY* array.

Discussion

Attempts to understand the nature and dynamics of Y-chromosomal rearrangements through combined physical mapping and haplotyping have been ongoing for a decade (43), but the current availability of the sequence (3), a robust phylogeny (41), and fine resolution of haplotypes (42) now permits detailed characterisation to be carried out. Our analysis of Y chromosomes ascertained by deletion of *AMELY* shows that recurrent 3.0-3.8-Mb deletions on Yp exist and persist within human populations.

Mechanisms and structural aspects

Deletion mapping and haplotyping analysis confirms that NAHR between *TSPY* copies is responsible for the deletion in the first two males described (17), and demonstrates that Class I deletions such as this recur, with seven independent instances observed. In principle the mutation rate can be calculated given an estimate of the number of generations encompassed within the phylogeny relating the sampled chromosomes (9, 44). Such an estimate, 52,000 generations, has been calculated for a set of 47 chromosomes representing the major branches of the phylogeny (9). For our Himalayan (26, 45), Southern Indian (46) and Spanish (22) samples, a total of 2471 chromosomes were analysed and three independent Class I deletion events observed. Time elapsed within the phylogeny relating these chromosomes is difficult to estimate; the sample size is large, but many chromosomes are closely related, and the deepest-rooting haplogroups (A and B) are absent, so an estimate of 100,000 generations is likely to be generous, in which case a conservative minimum mutation rate would be of the order of 10^{-5} per generation. The ease of detection of deletions through the amelogenin sex-test and the propensity to report deletions among the forensic community has led to a strong ascertainment bias, and the inclusion in our

collection of three examples of non-recurrent and rare deletions, possibly caused by non-homologous mechanisms.

Deletion mapping was undertaken in the knowledge that a 3.6-Mb segment of Yp exists in two different orientations (11, 12) due to an inversion polymorphism sponsored by the IR3 repeats (3). In the reference sequence, the orientations of the two *TSPY* loci are the same, so the arrangement is permissive for recombination and resulting Class I deletion (*Figure 4a*). However, because the single copy of *TSPY* lies distally in the distal IR3 element, most inversion events acting on chromosomes resembling the reference sequence are likely to preserve the distal copy's orientation while inverting the *TSPY* major array, resulting in a structure that is non-permissive for deletion (*Figure 4b*). Only inversions with breakpoints towards the outer ends of the IR3 elements will preserve the permissive orientation (*Figure 4c*). We can therefore conclude either that the seven progenitors of the Class I deletions possessed the same orientation of the inversion as the reference sequence, or that the inversion breakpoints are clustered towards the peripheries of the IR3 element pair. Inversion is recurrent, with at least 12 events observed among 47 Y chromosomes across the Y phylogeny (9), although one specific breakpoint has been shown to be monophyletic, and therefore a probable unique event (Hurles *et al.*, in preparation). Notably, the reference sequence orientation is seen in Class-I-containing haplogroups H and D, as well as R1b3, while the opposite orientation is seen in hgI and J2e (9). Further work is required to understand the molecular details of these inversions.

As with the other examples of repeat-mediated deletions on the Y chromosome (9, 47), we expect the reciprocal *TSPY*-mediated duplications to occur. Males carrying such duplications will have a duplicated copy of the Y-STR DYS458, which may be identified as a 'double-allele' when Y-STR profiling kits are applied. Two DYS458-duplication cases exist in the HGDP-CEPH diversity panel (HGDP388

and HGDP1153; data not shown), but semi-quantitative analysis of *AMELX/Y* signals provide no evidence of a duplication of *AMELY* in these particular cases.

Genes and selection

PRKY, *TBL1Y* and *AMELY* all belong to the so-called 'X-degenerate' class of Y-chromosomal genes, which were once identical in sequence to their X-chromosomal orthologues, but have degenerated in sequence since recombination was suppressed between the regions of the proto-sex-chromosomes in which they resided. However, this degeneration has not been ongoing in recent human evolution, since sequence comparisons of chimpanzee and human orthologues of X-degenerate genes show clear evidence of purifying selection in the human lineage (48). The genes commonly lost in *AMELY* deletions are, therefore, likely to be of functional significance.

The expansion of deletion lineages such as that in J2e1*/M241, and direct evidence demonstrating heritability in some cases, including 121/04, 710/01 (25), 37114 (19), VIC-B (28) and CM2, indicates that there cannot be a strong general selective disadvantage or negative effect on fertility. However, there may be more subtle effects resulting from the absence of genes on Yp. Because most of our samples were ascertained in population studies, direct phenotypic information is unavailable, so inferences in most cases must be made indirectly.

AMELY itself encodes a protein expressed in developing tooth-buds (49), and has an X-homologue (50), mutation of which causes X-linked amelogenesis imperfecta (51). However, *AMELY* is expressed at only 10% of the level of *AMELX*, and examination of two cases of *AMELY* deletion revealed apparently normal teeth (25), suggesting little or no effect of the deletion on enamel formation. In Class I deletions *PRKY*, a ubiquitously expressed serine-threonine protein kinase, and *TBL1Y*

(*Transducin [beta]-like 1 protein Y*), which is expressed in prostate and foetal brain (3) are also absent. The function of these genes is unknown, and again, examination of the phenotypes of *AMELY* deletion males has revealed no obvious phenotypic consequences of their absence (25). However, both have X-homologues, which may give clues to their functions. *PRKX* has been proposed to regulate tubulogenesis in the kidney (52), and *TBLIX* plays an essential role in specific nuclear receptor-mediated gene activation events (53), its deletion being implicated in late-onset sensorineural deafness (54). If the X and Y copies of these genes indeed share identical or similar functions, then males with deletions would be haploinsufficient for their products; interestingly, 45,X Turner Syndrome females, who are haploinsufficient for all XY-homologous genes, show both an increased incidence of renal abnormalities and late-onset sensorineural deafness (55). Careful monitoring of these phenotypes in *AMELY* deletion males would be worthwhile.

Class I deletions are caused by *TSPY*-mediated recombination, and the deletion event itself is accompanied by a reduction in the copy number of the repeat containing the *TSPY* gene. Maintenance of a minimum copy number through selection is suggested by the evolutionary conservation of multiple copies of the gene on the Y chromosomes of other mammals (56-59), and the limited degree of copy number polymorphism observed in two studies of human Y chromosomes (9, 60) (respectively, range: 18-47 or 23-64 copies; median 29 or 32 copies; sample size 89 or 47). In the one case we have been able to examine, within haplogroup J2e1*/M241, the reduction in number results in retention of at least 20 copies of *TSPY*. It would be of interest to determine copy numbers in further examples of independent Class I deletions.

In addition to the genes discussed above, the non-Class I deletions all lack *PCDH11Y* (*Protocadherin 11, Y-linked*), expressed in brain. This gene lies in the 'X-

transposed' segment of the Y chromosome (3) which has arisen since our divergence from the human-chimpanzee common ancestor, and therefore represents a human-specific genetic novelty. Protocadherins are important in cell-cell interactions within the central nervous system, and differences in structure and expression of the X and Y copies, and accelerated evolution within the human lineage (61), have led to the nomination of *PCDH11X/Y* as candidates for sex-specific differences in human brain phenotypes (62). Unfortunately information on cognitive ability in the males lacking *PCDH11Y* is unavailable. The single Class III case also lacks the testis-expressed *TGIF2LY* (*Transforming Growth Factor-Beta-Induced Factor 2-Like, Y-Linked* (63)), about which little is known; this male is a singleton, with no phenotypic information, so we cannot exclude the possibility that absence of *TGIF2LY* has a deleterious effect on spermatogenesis.

Deletion of the *AMELY* region has one selectively beneficial effect: the descendants of deleted males are protected against sex-reversal caused by the commonest class of XY-translocation. The deleted segment in all of the males encompasses an XY-homologous region, around the *PRKY* gene, that (in one orientation (64)) sponsors XY-translocations which cause XX maleness (65).

Forensic implications

The geographical origins of males carrying the common hgJ2e1*/M241 Class I deletion are mainly in the Indian subcontinent; nine were from Nepal, ten from a Malaysian population of Indian origin (29), two from Sri Lanka (17), and two from the Maldives Islands in the Indian Ocean, though there are geographical outliers in Afghanistan, Morocco and Australia. Interestingly, a second cluster of Class I deletions, in hgH(xH2), also appears to have an Indian origin. The relatively high frequency of *AMELY* deletion chromosomes in the Indian subcontinent (~2%) and

around the Indian Ocean seems likely to be a result of genetic drift. However, it has practical implications for the massive forensic identification effort following the South Asian tsunami disaster of 2004, since it is in this part of the world that deletions are most likely to introduce confusion through the misassignment of sex based on DNA evidence. In cases where other evidence is lacking or ambiguous, confirmation using an additional test, such as presence/absence of *SRY* (17), is necessary.

Materials and Methods

DNA samples

Male HGDP644 is from the CEPH-HGDP diversity panel (66), and was identified as *AMELY*-deleted during profiling of the panel using the Identifiler autosomal STR kit (Applied Biosystems). Other DNA samples were from collections of the authors, and obtained with appropriate informed consent. CM2 was ascertained through a positive paternity test, demonstrating that he was fertile, and HE4 was ascertained during a study of 100 males from Southern India (46). Some samples were subjected to whole genome amplification (67) using the Genomiphi kit (GE Healthcare) before analysis.

Deletion mapping

Y-specific STSs (primer sequences available from the literature (3, 36, 37)) were amplified by PCR and analysed by agarose gel electrophoresis. Some new or adapted STSs were designed to amplify Y-specific sequences in blocks of XY-homology (*Supplementary Table 1*), by targeting the 3' ends of primers at Y-specific bases identified in sequence alignments. An STS was considered to be deleted when absent in the presence of a larger Y-specific control amplicon coamplified in the same PCR reaction. The control sequence generally used was an amplicon from the *SRY* gene region (715 bp, primers 5'-AAT CGG GTA ACA TTG GCT ACA-3', and 5'-AGG CTT AAA AGT TAA TAG GCC A-3'). PCR was carried out in a Tetrad Thermocycler (MJR) using 5-10 ng template DNA, the buffer of Jeffreys et al. (68) and 0.1u *Taq* DNA polymerase (ABgene). PCR conditions were generally: 94°C 30s, 60°C 30s, 70°C 30s, for 33 cycles, though annealing temperature was varied to ensure Y-specificity in some cases. Not all STSs were mapped in all samples, because the

amount of DNA was limiting. STS positions were obtained from the UCSC Genome Browser (<http://genome.ucsc.edu/>), using Build 36.1 of the reference sequence.

Estimation of TSPY copy number

Primers were designed to amplify a 78-bp region of the *TSPY* repeat and amplification efficiency was checked with a DNA dilution series using a region of the diploid *PMP22* gene as a control (69). Four *TSPY* and four *PMP22* replicates were set up for each sample in 96-well plates, each in a total volume of 25 μ l, containing 1 x SYBR Green master mix (Applied Biosystems), 25ng template DNA, 300nM primer TSPYqF (5'-GAG TCC CCT GAC AGA TCC TAT GTA A-3') and primer TSPYqR (5'-CTT CAG GTG GCT TCA TCC TCT T-3'). Cycling was performed on an Applied Biosystems 7500 quantitative PCR machine using cycling conditions: 95°C for 10 minutes followed by 40 rounds of 95°C for 15 seconds, 62°C for 1 minute. All DNAs analysed had not been subject to whole genome amplification.

Difference in threshold cycle ($\bullet C_t$) was calculated for each individual by subtracting mean *PMP22* C_t from mean *TSPY* C_t . Using one individual as a reference, values were converted to $\bullet\bullet C_t$, from which relative copy number was calculated (70). Absolute copy numbers of *TSPY* genes were estimated with reference to DNA samples in which copy number was known, from size measurement of the hybridising *Xba*I fragment in pulsed-field gel analysis (71). Comparison of PCR-based and pulsed-field methods suggest that the former is accurate to ± 1 *TSPY* repeat unit.

Y chromosome haplotyping

Binary markers were typed in a hierarchical fashion using the SNaPshot minisequencing protocol (Applied Biosystems) on an ABI3100 capillary electrophoresis apparatus (Applied Biosystems). Amplification primers and SNaPshot

primers were based on ones published previously (72, 73), with additional primers based on published sequences (41).

Twenty-six Y-STRs (DYS19, *DYS385a/b*, *DYS388*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS425*, *DYS426*, *DYS434*, *DYS435*, *DYS436*, *DYS437*, *DYS438*, *DYS439*, *DYS447*, *DYS448*, *DYS460*, *DYS461*, *DYS462*, *YCAII-sY59/b*, and *Y-GATA-H4.1*) were typed in a 20-plex (74) and an additional 14-plex (26) that incorporates the amelogenin sex-test (14). PCR products were resolved on an ABI3100 capillary electrophoresis apparatus (Applied Biosystems), and analysed using GeneMapper software (Applied Biosystems). Allele nomenclature was as described (26).

Reputer analysis

Perfect direct repeats within deletion breakpoint intervals were identified using the Reputer program (38), at <http://bibiserv.techfak.uni-bielefeld.de/reputer/>.

Y-STR network construction and dating

A weighted median-joining network (75) was constructed from Y-STR haplotypes using Network 4.0 (<http://www.fluxus-engineering.com/sharenet.htm>). The network contains all the most parsimonious trees relating the haplotypes, and allows unobserved haplotypes to represent nodes (branch points). Weighting (76) was used to remove some reticulations (closed structures) within the network by taking into account the range of different mutation rates of the Y-STRs, reflected indirectly by their allele length variances among all deletion chromosomes. Thus, fast-mutating loci (high variance) are given lower weight in making links between haplotypes than are slow-mutating loci (low variance). TMRCA of the hgJ2e1*/M241 cluster was

estimated within Network from the rho statistic, using a 25-year generation time and a mean per-locus, per-generation mutation rate of 6.9×10^{-4} (77).

Acknowledgements

We thank all DNA donors, and Aleksander Bagdonavicius (Path Centre Forensic Biology, WA, Australia), for provision of DNA samples. M.A.J. is a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant number 057559); S.M.A., A.C.L., Y.X. and C.T.S. were supported by the Wellcome Trust. E.J.P. was supported by the Arts and Humanities Research Council and the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409, within the framework of the European Science Foundation EUROCORES programme “The Origin of Man, Language and Languages”. T.K. and P.dK. were supported by the NWO (Netherlands Organisation for Scientific Research) project 231-70-001 within the same EUROCORES programme.

References

1. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003-1007.
2. Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E. *et al.* (2006) Copy number variation: new insights in genome diversity. *Genome Res.*, **16**, 949-961.
3. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, R., Ali, J., Bieri, T. *et al.* (2003) The male-specific region of the human Y chromosome: a mosaic of discrete sequence classes. *Nature*, **423**, 825-837.
4. Bobrow, M., Pearson, P.L., Pike, M.C. and El-Alfi, O.S. (1971) Length variation in the quinacrine binding segment of human Y chromosomes of different sizes. *Cytogenet.*, **10**, 190-198.
5. Verma, R.S., Dosik, H., Scharf, T. and Lubs, H.A. (1978) Length heteromorphisms of fluorescent (f) and non-fluorescent (nf) segments of human Y chromosome: classification, frequencies, and incidence in normal Caucasians. *J. Med. Genet.*, **15**, 277-281.
6. Verma, R.S., Rodriguez, J. and Dosik, H. (1982) The clinical significance of pericentric inversion of the human Y chromosome: a rare "third" type of heteromorphism. *J. Hered.*, **73**, 236-238.
7. Bernstein, R., Wadee, A., Rosendorff, J., Wessels, A. and Jenkins, T. (1986) Inverted Y chromosome polymorphism in the Gujerati Muslim Indian population of South Africa. *Hum. Genet.*, **74**, 223-229.

8. Schmid, M., Haaf, T., Solleder, E., Schempp, W., Leipoldt, M. and Heilbronner, H. (1984) Satellited Y chromosomes: structure, origin and clinical significance. *Hum. Genet.*, **67**, 72-85.
9. Repping, S., van Daalen, S.K., Brown, L.G., Korver, C.M., Lange, J., Marszalek, J.D., Pyntikova, T., van der Veen, F., Skaletsky, H., Page, D.C. *et al.* (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat. Genet.*, **38**, 463-467.
10. Tyler-Smith, C., Taylor, L. and Müller, U. (1988) Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *J. Mol. Biol.*, **203**, 837-848.
11. Affara, N.A., Ferguson-Smith, M.A., Tolmie, J., Kwok, K., Mitchell, M., Jamieson, D., Cooke, A. and Florentin, L. (1986) Variable transfer of Y-specific sequences in XX males. *Nucl. Acids Res.*, **14**, 5375-5387.
12. Page, D.C. (1986) Sex reversal: deletion mapping the male-determining function of the human Y chromosome. *Cold Spring Harb. Symp. Quant. Biol.*, **51**, 229-235.
13. Zhu, B., Sun, Q.W., Lu, Y.C., Sun, M.M., Wang, L.J. and Huang, X.H. (2005) Prenatal fetal sex diagnosis by detecting amelogenin gene in maternal plasma. *Prenat. Diagn.*, **25**, 577-581.
14. Sullivan, K.M., Mannucci, A., Kimpton, C.P. and Gill, P. (1993) A rapid and quantitative DNA sex test - fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques*, **15**, 636.
15. Stone, A.C., Milner, G.R., Pääbo, S. and Stoneking, M. (1996) Sex determination of ancient human skeletons using DNA. *Am. J. Phys. Anthropol.*, **99**, 231-238.

16. Faerman, M., Kahila, G., Smith, P., Greenblatt, C., Stager, L., Filon, D. and Oppenheim, A. (1997) DNA analysis reveals the sex of infanticide victims. *Nature*, **385**, 212-213.
17. Santos, F.R., Pandya, A. and Tyler-Smith, C. (1998) Reliability of DNA-based sex tests. *Nat. Genet.*, **18**, 103.
18. Roffey, P.E., Eckhoff, C.I. and Kuhl, J.L. (2000) A rare mutation in the amelogenin gene and its potential investigative ramifications. *J. Forensic Sci.*, **45**, 1016-1019.
19. Henke, J., Henke, L., Chatthopadhyay, P., Kayser, M., Dülmer, M., Cleef, S., Pöche, H. and Felske-Zech, H. (2001) Application of Y-chromosomal STR haplotypes to forensic genetics. *Croat. Med. J.*, **42**, 292-297.
20. Thangaraj, K., Reddy, A.G. and Singh, L. (2002) Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? *Int. J. Legal Med.*, **116**, 121-123.
21. Steinlechner, M., Berger, B., Niederstatter, H. and Parson, W. (2002) Rare failures in the amelogenin sex test. *Int. J. Legal Med.*, **116**, 117-120.
22. Bosch, E., Lee, A.C., Calafell, F., Arroyo, E., Henneman, P., de Knijff, P. and Jobling, M.A. (2002) High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forens. Sci. Int.*, **125**, 42-51.
23. Chang, Y.M., Burgoyne, L.A. and Both, K. (2003) Higher failures of amelogenin sex test in an Indian population group. *J. Forensic Sci.*, **48**, 1309-1313.
24. Michael, A. and Brauner, P. (2004) Erroneous gender identification by the amelogenin sex test. *J. Forensic Sci.*, **49**, 258-259.
25. Lattanzi, W., Di Giacomo, M.C., Lenato, G.M., Chimienti, G., Voglino, G., Resta, N., Pepe, G. and Guanti, G. (2005) A large interstitial deletion

- encompassing the amelogenin gene on the short arm of the Y chromosome. *Hum. Genet.*, **116**, 395-401.
26. Parkin, E.J., Kraayenbrink, T., van Driem, G.L., Tshering, K., de Knijff, P. and Jobling, M.A. (2006) 26-locus Y-STR typing in a Bhutanese population sample. *Forens. Sci. Int.*, **161**, 1-7.
 27. Kashyap, V.K., Sahoo, S., Sitalaximi, T. and Trivedi, R. (2006) Deletions in the Y-derived amelogenin gene fragment in the Indian population. *BMC Med. Genet.*, **7**, 37.
 28. Mitchell, R.J., Kreskas, M., Baxter, E., Buffalino, L. and Van Oorschot, R.A. (2006) An investigation of sequence deletions of amelogenin (AMELY), a Y-chromosome locus commonly used for gender determination. *Ann. Hum. Biol.*, **33**, 227-240.
 29. Chang, Y.M., Perumal, R., Keat, P.Y., Yong, R.Y., Kuehn, D.L. and Burgoyne, L. (2006) A distinct Y-STR haplotype for Amelogenin negative males characterized by a large Y(p)11.2 (DYS458-MSY1-AMEL-Y) deletion. *Forens. Sci. Int.*, **in press**.
 30. Cadenas, A.M., Regueiro, M., Gayden, T., Singh, N., Zhiotovskiy, L.A., Underhill, P.A. and Herrera, R.J. (2006) Male amelogenin dropouts: phylogenetic context, origins and implications. *Forens. Sci. Int.*, **in press**.
 31. Sun, C., Skaletsky, H., Rozen, S., Gromoll, J., Nieschlag, E., Oates, R. and Page, D.C. (2000) Deletion of *azoospermia factor a* (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum. Mol. Genet.*, **9**, 2291-2296.
 32. Kamp, C., Hirschmann, P., Voss, H., Huellen, K. and Vogt, P.H. (2000) 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.
33. Blanco, P., Shlumukova, M., Sargent, C.A., Jobling, M.A., Affara, N. and Hurles, M.E. (2000) Divergent outcomes of intra-chromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.*, **37**, 752-758.
34. Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C. and Rozen, S. (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am. J. Hum. Genet.*, **71**, 906-922.
35. Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S. *et al.* (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.*, **29**, 279-286.
36. Vollrath, D., Foote, S., Hilton, A., Brown, L.G., Beer-Romero, P., Bogan, J.S. and Page, D.C. (1992) The human Y chromosome - a 43-interval map based on naturally-occurring deletions. *Science*, **258**, 52-59.
37. Tilford, C.A., Kuroda-Kawaguchi, T., Skaletsky, H., Rozen, S., Brown, L.G., Rosenberg, M., McPherson, J.D., Wylie, K., Sekhon, M., Kucaba, T.A. *et al.* (2001) A physical map of the human Y chromosome. *Nature*, **409**, 943-945.
38. Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J. and Giegerich, R. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucl. Acids Res.*, **29**, 4633-4642.
39. Lukacsovich, T. and Waldman, A.S. (1999) Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics*, **151**, 1559-1568.

40. Shaw, C.J. and Lupski, J.R. (2005) Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum. Genet.*, **116**, 1-7.
41. Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.*, **12**, 339-348.
42. Jobling, M.A. and Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.*, **4**, 598-612.
43. Jobling, M.A., Samara, V., Pandya, A., Fretwell, N., Bernasconi, B., Mitchell, R.J., Gerelsaikh, T., Dashnyam, B., Sajantila, A., Salo, P.J. *et al.* (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Hum. Mol. Genet.*, **5**, 1767-1775.
44. Bosch, E., Hurles, M.E., Navarro, A. and Jobling, M.A. (2004) Dynamics of a human interparalog gene conversion hotspot. *Genome Res.*, **14**, 835-844.
45. Parkin, E.J., Kraayenbrink, T., Opgenort, J.R., van Driem, G.L., Tuladhar, N.M., de Knijff, P. and Jobling, M.A. (2006) Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: Isolation and drift in the Himalayas. *Forens. Sci. Int.*, **in press**.
46. Lee, A.C., Kamalam, A., Adams, S.M. and Jobling, M.A. (2004) Molecular evidence for absence of Y-linkage of the Hairy Ears trait. *Eur. J. Hum. Genet.*, **12**, 1077-1079.
47. Bosch, E. and Jobling, M.A. (2003) Duplications of the *AZFa* region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Hum. Mol. Genet.*, **12**, 341-347.
48. Hughes, J.F., Skaletsky, H., Pyntikova, T., Minx, P.J., Graves, T., Rozen, S., Wilson, R.K. and Page, D.C. (2005) Conservation of Y-linked genes during

- human evolution revealed by comparative sequencing in chimpanzee. *Nature*, **437**, 100-103.
49. Salido, E.C., Yen, P.H., Koprivnikar, K., Yu, L.C. and Shapiro, L.J. (1992) The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am. J. Hum. Genet.*, **50**, 303-316.
50. Lau, E.C., Mohandas, T.K., Shapiro, L.J., Slavkin, H.C. and Snead, M.L. (1989) Human and mouse amelogenin gene loci are on the sex chromosomes. *Genomics*, **6**, 162-168.
51. Lagerstrom, M., Dahl, N., Nakahori, Y., Nakagome, Y., Backman, B., Landegren, U. and Pettersson, U. (1991) A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). *Genomics*, **10**, 971-975.
52. Li, X., Li, H.P., Amsler, K., Hyink, D., Wilson, P.D. and Burrow, C.R. (2002) PRKX, a phylogenetically and functionally distinct cAMP-dependent protein kinase, activates renal epithelial cell migration and morphogenesis. *Proc. Natl. Acad. Sci. USA*, **99**, 9260-9265.
53. Perissi, V., Aggarwal, A., Glass, C.K., Rose, D.W. and Rosenfeld, M.G. (2004) A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell*, **116**, 511-526.
54. Bassi, M.T., Ramesar, R.S., Caciotti, B., Winship, I.M., De Grandi, A., Riboni, M., Townes, P.L., Beighton, P., Ballabio, A. and Borsani, G. (1999) X-linked late-onset sensorineural deafness caused by a deletion involving OA1 and a novel gene containing WD-40 repeats. *Am. J. Hum. Genet.*, **64**, 1604-1616.

55. Elsheikh, M., Dunger, D.B., Conway, G.S. and Wass, J.A. (2002) Turner's syndrome in adulthood. *Endocr. Rev.*, **23**, 120-140.
56. Guttenbach, M., Müller, U. and Schmid, M. (1992) A human moderately repeated Y-specific DNA sequence is evolutionarily conserved in the Y chromosome of the great apes. *Genomics*, **13**, 363-367.
57. Jakubiczka, S., Schnieders, F. and Schmidtke, J. (1993) A bovine homologue of the human TSPY gene. *Genomics*, **17**, 732-735.
58. Raudsepp, T., Santani, A., Wallner, B., Kata, S.R., Ren, C., Zhang, H.B., Womack, J.E., Skow, L.C. and Chowdhary, B.P. (2004) A detailed physical map of the horse Y chromosome. *Proc. Natl. Acad. Sci. USA*, **101**, 9321-9326.
59. Murphy, W.J., Pearks Wilkerson, A.J., Raudsepp, T., Agarwala, R., Schaffer, A.A., Stanyon, R. and Chowdhary, B.P. (2006) Novel gene acquisition on carnivore Y chromosomes. *PLoS Genet.*, **2**, e43.
60. Mathias, N., Bayés, M. and Tyler-Smith, C. (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum. Mol. Genet.*, **3**, 115-123.
61. Williams, N.A., Close, J.P., Giouzei, M. and Crow, T.J. (2006) Accelerated evolution of *Protocadherin11X/Y*: a candidate gene-pair for cerebral asymmetry and language. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **141**, 623-633.
62. Blanco, P., Sargent, C.A., Boucher, C., Mitchell, M. and Affara, N.A. (2000) Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain. *Mamm. Genome*, **11**, 906-914.
63. Blanco-Arias, P., Sargent, C.A. and Affara, N.A. (2002) The human-specific Yp11.2/Xq21.3 homology block encodes a potentially functional testis-specific TGIF-like retroposon. *Mamm. Genome*, **13**, 463-468.

64. Jobling, M.A., Williams, G., Schiebel, K., Pandya, A., McElreavey, K., Salas, L., Rappold, G.A., Affara, N.A. and Tyler-Smith, C. (1998) A selective difference between human Y-chromosomal DNA haplotypes. *Curr. Biol.*, **8**, 1391-1394.
65. Schiebel, K., Winkelmann, M., Mertz, A., Xu, X., Page, D.C., Weil, D., Petit, C. and Rappold, G.A. (1997) Abnormal XY interchange between a novel isolated protein kinase gene, *PRKY*, and its homologue, *PRKX*, accounts for one third of all (Y+)XX males and (Y-)XY females. *Hum. Mol. Genet.*, **6**, 1985-1989.
66. Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A. *et al.* (2002) A human genome diversity cell line panel. *Science*, **296**, 261-262.
67. Dean, F.B., Hosono, S., Fang, L.H., Wu, X.H., Faruqi, A.F., Bray-Ward, P., Sun, Z.Y., Zong, Q.L., Du, Y.F., Du, J. *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA*, **99**, 5261-5266.
68. Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell*, **60**, 473-485.
69. McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J. *et al.* (2006) Common deletion polymorphisms in the human genome. *Nat. Genet.*, **38**, 9-11.
70. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods*, **25**, 402-408.

71. Oakey, R. and Tyler-Smith, C. (1990) Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. *Genomics*, **7**, 325-330.
72. Paracchini, S., Arredi, B., Chalk, R. and Tyler-Smith, C. (2002) Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucl. Acids Res.*, **30**, e27.
73. Bosch, E., Calafell, F., González-Neira, A., Flaiz, C., Mateu, E., Scheil, H.-G., Huckenbeck, W., Efremovska, L., Mikerezi, I., Xirotiris, N. *et al.* (2006) Male and female lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann. Hum. Genet.*, **70**, 459-487.
74. Butler, J.M., Schoske, R., Vallone, P.M., Kline, M.C., Redd, A.J. and Hammer, M.F. (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forens. Sci. Int.*, **129**, 10-24.
75. Bandelt, H.-J., Forster, P. and Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, **16**, 37-48.
76. Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C. and Mehdi, S.Q. (2002) Y-chromosomal DNA variation in Pakistan. *Am. J. Hum. Genet.*, **70**, 1107-1124.
77. Zhivotovsky, L.A., Underhill, P.A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G. *et al.* (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.*, **74**, 50-61.

Figure legends

Figure 1: Schematic representation of deletion mapping data.

Black bars indicate regions of Yp observed or inferred to be present, and grey bars indicate regions of uncertainty within deletion breakpoint intervals. ‘+’: marker present; ‘-’: marker absent; ‘I-s’: Class I deletions also lacking sY59; ‘Aus.’: Australian. Genes are shown below. The scale in Mb shows distance from the Yp telomere, according to the March 2006 (Build 36.1) genome assembly. Haplogroups are given for completeness (see text and *Figure 3a*), and J2e1* indicates J2e1*/M241. ‘Ref’ – reference; those without numbers are from this study.

Figure 2: Haplogroups and Y-STR haplotypes of *AMELY* deletion chromosomes.

a) Haplogroups (Hg) found in deletion chromosomes, indicated by coloured boxes. The phylogeny shows key binary markers typed to define haplogroups.

b) Median-joining network of Y-STR haplotypes. Circles represent haplotypes, with area proportional to frequency, and coloured by haplogroup according to (a); different classes of deletions are indicated by circle outlines, as in the key below the phylogeny. Lines between the circles indicate mutational differences between Y-STR haplotypes, with lengths proportional to the number of inferred mutational steps. The 20 Y-STRs used for construction were DYS19, DYS388, DYS389I, DYS389II-I, DYS390, DYS391, DYS392, DYS393, DYS434, DYS435, DYS436, DYS437, DYS438, DYS439, DYS447, DYS448, DYS460, DYS461, DYS462, and YGATA-H4.1. The partial allele (26.2) at DYS447 in male N5 was considered as allele 26 for network construction.

Figure 3: *TSPY* copy number in *AMELY*-deleted and non-deleted chromosomes within hgJ2e1*/M241.

For deletion chromosomes, names of samples are given within the squares. Grey shading indicates the Class I chromosome lacking sY59.

Figure 4: Schematic illustration of the influence of IR3-mediated inversion on *TSPY*-mediated deletion.

Some structures are permissive for the deletion; others are not. Diagram not to scale.

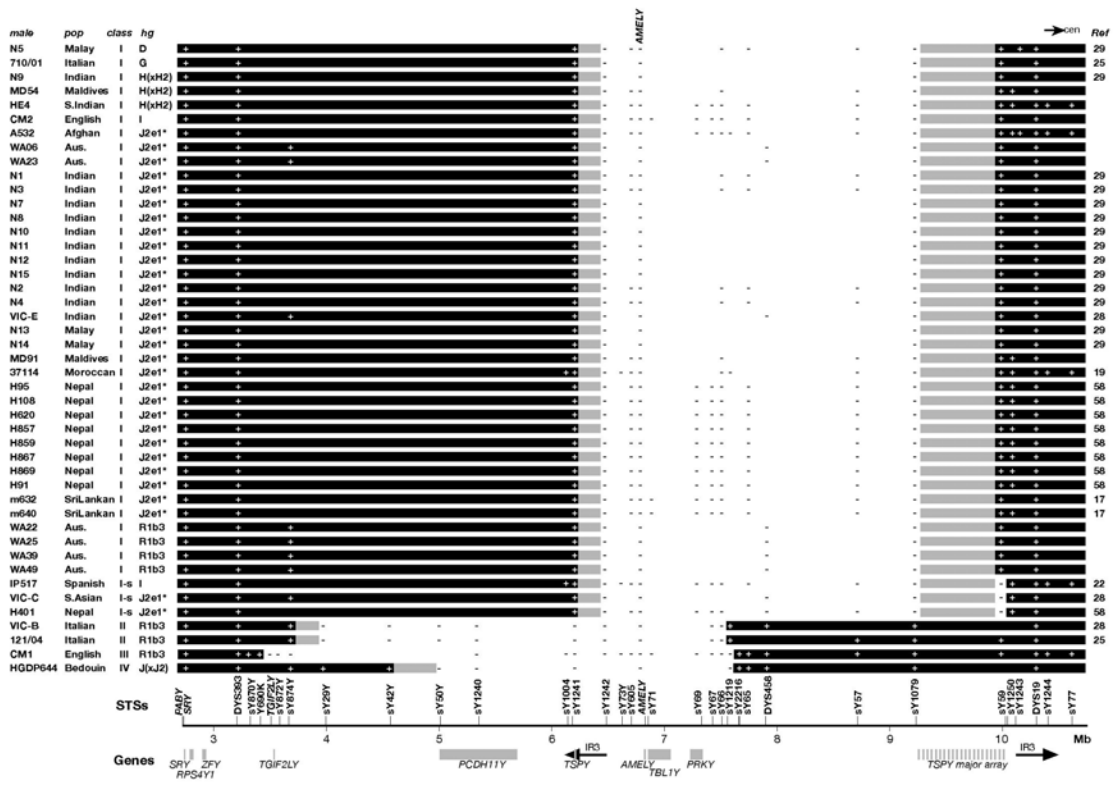


Figure 1

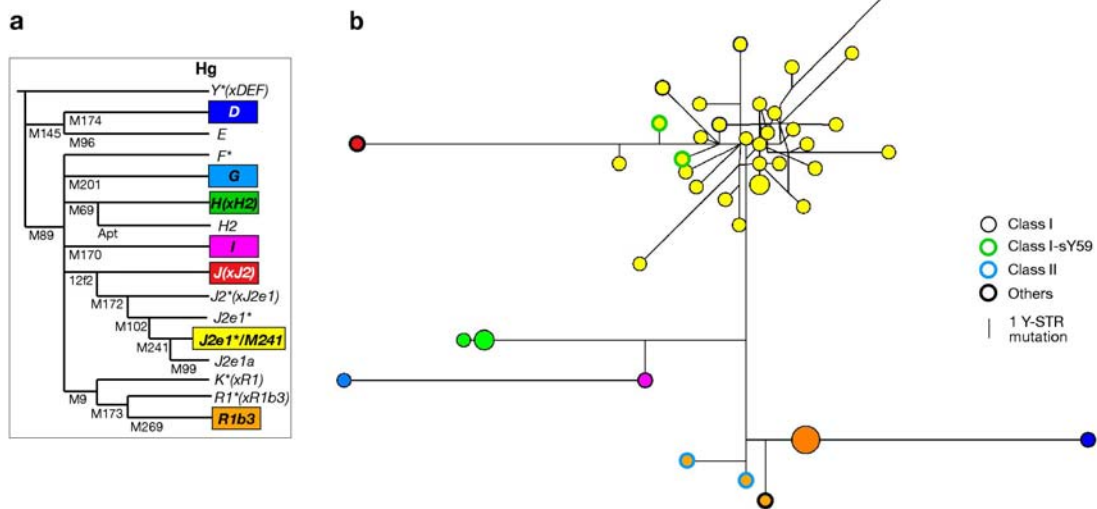


Figure 2

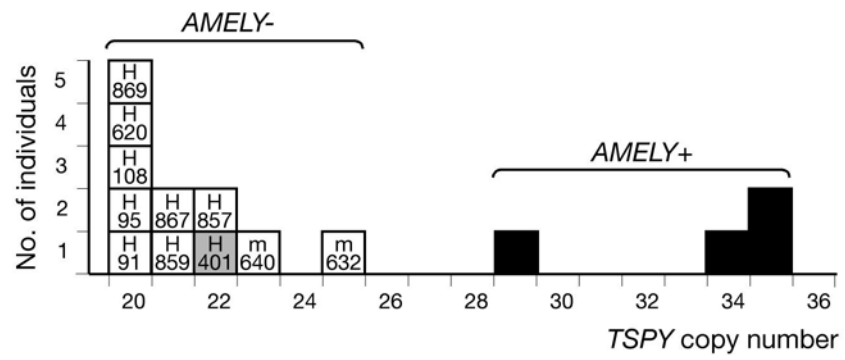


Figure 3

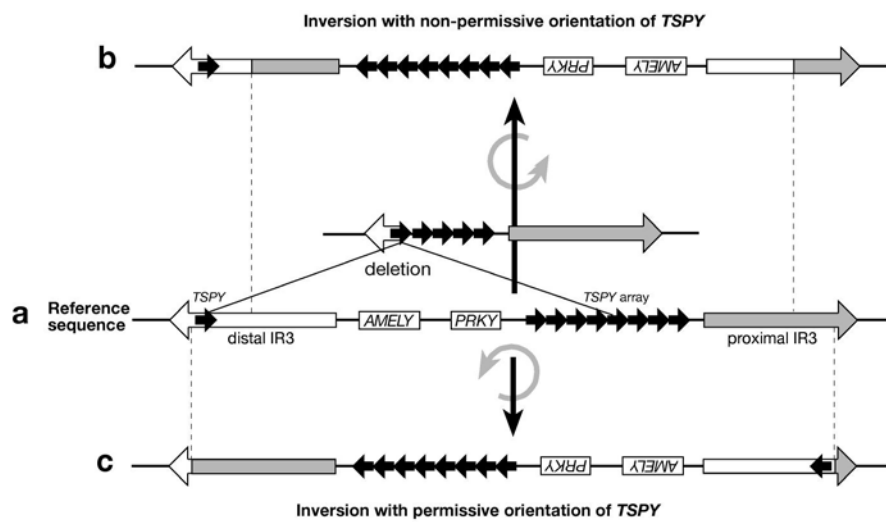


Figure 4