

SHORT REPORT

# Genealogy and genes: tracing the founding fathers of Tristan da Cunha

Himla Soodyall<sup>\*,1,2</sup>, Almut Nebel<sup>1,2</sup>, Bharti Morar<sup>1</sup> and Trefor Jenkins<sup>1</sup>

<sup>1</sup>MRC/NHLS/Wits Human Genomic Diversity and Disease Research Unit, Division of Human Genetics, School of Pathology, National Health Laboratory Service and University of the Witwatersrand, Johannesburg 2000, South Africa

The island population of Tristan da Cunha has a well-documented genealogy that dates to its first permanent settlement in 1816. The current population is thought to have descended from only seven females and eight males. Today, there are seven family names in use, corresponding to the number of founding fathers with present-day male descendants. Y chromosome polymorphisms have previously been shown to be reliable tools for tracing patrilineal genealogies. Here, we studied Y chromosome polymorphisms in a sample from Tristan da Cunha together with genealogical records to (i) infer the haplotypes of the seven founders and (ii) test if the Y chromosome transmission is consistent with the documented patrilineal history of the island community. We observed nine Y chromosome haplotypes of which seven could be traced to the known ancestors. Of the two additional lineages, one probably evolved from a founder haplotype due to a single-step microsatellite mutation, while the other had an obvious non-island origin. Its introduction, however, is not reflected in the records. Four more instances of non-paternity were identified, with the 'new' chromosomes matching other island haplotypes. The Y chromosome data presented here question the validity of some of the genealogical documentation and emphasise the value of genetic studies in tracing ancestry.

*European Journal of Human Genetics* (2003) 11, 705–709. doi:10.1038/sj.ejhg.5201022

**Keywords:** Tristan da Cunha; genealogy; Y chromosome; bi-allelic and microsatellite polymorphisms

## Introduction

Genealogy is the discipline that traditionally deals with the reconstruction of family history. So far, the quest to trace ancestors and distant relatives has mainly relied on oral traditions, church records and other documents. The human genome constitutes an archive of our history and evolutionary past that can be explored for genealogical research. Genetic markers, like mitochondrial DNA (mtDNA) and Y chromosome polymorphisms, have been shown in historical investigations to be reliable tools for tracing matri- and patrilineal genealogies, respectively.<sup>1,2</sup>

Since hereditary family names are passed down in the same way as paternally inherited Y chromosomes, a surname should, within a genealogy, correlate with a Y chromosome haplotype.<sup>3,4</sup>

The island population of Tristan da Cunha is unique with regard to its well-documented genealogy that dates back to the first permanent settlement in the early 19th century.<sup>5</sup> The current population of 278 individuals<sup>6</sup> is thought to have descended from 15 ancestors, seven females and eight males who arrived on the island at various times between 1816 and 1908. The male founders were all of western European ancestry, originating from Scotland, England, Holland, the USA and Italy. Today, there are seven family names in use (Glass, Green, Hagan, Lavarello, Repetto, Rogers and Swain), corresponding to the number of founding fathers with present-day male descendants. In addition to these known donors, there is

\*Correspondence: Dr H Soodyall, Division of Human Genetics, School of Pathology, National Health Laboratory Service and University of the Witwatersrand, PO Box 1038, Johannesburg 2000, South Africa. Tel: +27-11-489 9208; Fax: +27-11-489 9209; E-mail: hxsood@global.co.za

<sup>2</sup>The first two authors contributed equally to the paper.

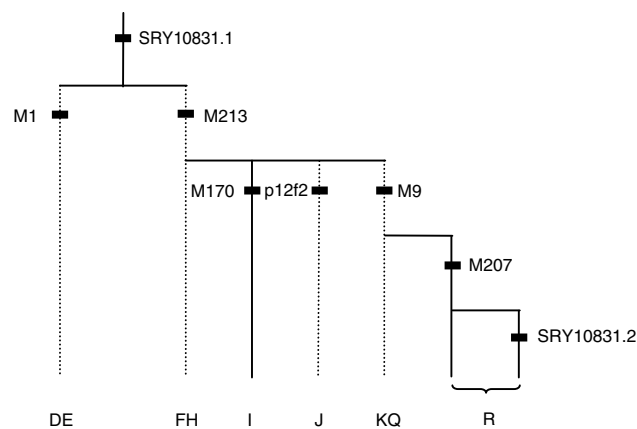
Received 18 February 2003; revised 24 March 2003; accepted 3 April 2003

also evidence for 'hidden ancestors' who supposedly contributed to the gene pool, but took their names with them upon leaving the island.<sup>7</sup>

In a previous study, the present-day mtDNA pool of the Tristan islanders was traced to five female ancestors.<sup>8</sup> These data identified inconsistencies in the number of mtDNA types and the written records. Although the historical documents mentioned two sister pairs among the founding females, mtDNA data showed support for only one pair of sisters. In the present study, we studied 13 Y chromosome polymorphisms in a sample from Tristan da Cunha together with the available genealogical information to (i) infer the Y chromosome haplotypes of the known male founders who brought their surnames to the island and (ii) test if the Y chromosome transmission is consistent with the documented patrilineal history of the island community.

### Subjects and methods

DNA samples were prepared from peripheral blood by standard phenol–chloroform procedures from 76 males in a sample collected on the island of Tristan da Cunha in 1982. This study was approved by the Committee for Research on Human Subjects at the University of Witwatersrand, Johannesburg, South Africa. Out of respect for the sensitivities of the present inhabitants of Tristan da Cunha, the observed haplotypes were neither associated with the family names nor with the origins of the founding fathers in order to avoid identification. DNA samples were screened for 13 Y chromosome-specific polymorphisms. The seven bi-allelic Y chromosome markers used here (Figure 1) have previously been shown to define haplogroups frequently found in European populations.<sup>9–11</sup>



**Figure 1** Y chromosome phylogeny of the bi-allelic polymorphisms used in the study (black bars). The three Y chromosome haplogroups found in the sample from Tristan da Cunha are indicated (solid lines). Haplogroups are designated according to the YCC nomenclature.<sup>12</sup>

Haplogroup designation follows the nomenclature proposed by the Y Chromosome Consortium (YCC).<sup>12</sup> The markers M1 (Y chromosome *Alu* insertion polymorphism 'YAP') and M9 were typed in all samples. The allelic states of the remaining polymorphisms were determined hierarchically based on the known Y chromosome genealogy (Figure 1).<sup>12</sup> M1, p12f2 and SRY10831 were screened as described.<sup>13,9,14</sup> M9,<sup>15</sup> M170, M207, M213<sup>16</sup> were typed using agarose gel-based PCR–RFLP assays. The protocols for PCR amplification and the detection of polymorphic markers are available from the authors upon request. The six microsatellite loci DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393<sup>17</sup> were amplified in two separate multiplex PCR reactions. The PCR conditions are available from the authors upon request. The fluorescently labelled PCR products were resolved on polyacrylamide gels in an automated ABI 377 DNA sequencer. The number of repeats at each locus was determined adopting the proposed nomenclature.<sup>18</sup> Haplotypes were defined by both bi-allelic and microsatellite polymorphisms. Microsatellite haplotypes were constructed using the number of repeats (alleles) in the order DYS19–DYS388–DYS390–DYS391–DYS392–DYS393.

### Results and discussion

We derived nine Y chromosome haplotypes of which seven could unambiguously be traced to the seven men who are known to have introduced surnames to the island (Table 1, haplotypes shown in bold). These seven haplotypes segregated either on a haplogroup R-M207 ( $\times$  SRY10831.2) or an I-M170 background. R-M207 ( $\times$  SRY10831.2) is mainly equivalent to R-M173 ( $\times$  M17),<sup>12</sup> which is the dominant Y chromosome lineage in western European populations where it reaches frequencies of 50% and above.<sup>10,11</sup> Haplogroup I-M170 is virtually confined to Europe. In western Europeans it has been observed at moderate frequencies (eg 22% in Dutch and 8% in Italians).<sup>10,11</sup> Hence, the presence of haplogroups R-M207 and I-M170 in the sample from Tristan da Cunha is consistent with the western European ancestry of the known founding fathers, as suggested by historical records.<sup>7</sup>

Moreover, two additional haplotypes were discovered (Table 1, haplotypes in families 3 and 4 shown in italics). The first one was found in family 3, in four individuals spanning three generations (Figure 2a). This haplotype could have evolved from one of the two haplotypes already present in families 3 and 5 (Table 1), because of single-step microsatellite mutations at either DYS391 or DYS392. Since the observed mutation rate at DYS392 is very low in European populations (0 in 415 meioses, 95% CI  $7.15 \times 10^{-3}$ ),<sup>19</sup> the mutation at DYS391 seems to be the most likely event that gave rise to the new haplotype

**Table 1** Y chromosome haplotypes found in the seven families on Tristan da Cunha

| Family <sup>a</sup> | No. of males <sup>b</sup> | Haplogroup <sup>c</sup>              | Haplotype <sup>d</sup>   |
|---------------------|---------------------------|--------------------------------------|--|
| 1                   | 5/5                       | <b>R-M207</b>                        | <b>15-12-25-10-14-13</b>   |
| 2                   | 3/3                       | <b>R-M207</b>                        | <b>14-12-24-11-13-13</b>   |
| 3                   | <b>9/13</b><br>4/13       | <b>R-M207</b><br><i>R-M207</i>       | <b>14-12-23-11-13-13</b><br><i>14-12-23-10-13-13<sup>e</sup></i> |
| 4                   | <b>8/9</b><br>1/9         | <b>R-M207</b><br><i>R-SRY10831.2</i> | <b>14-12-24-10-13-14</b><br><i>16-12-25-10-11-13<sup>f</sup></i> |
| 5                   | <b>16/19</b><br>3/19      | <b>R-M207</b><br><i>I-M170</i>       | <b>14-12-23-10-14-13</b><br><i>14-14-22-10-11-13<sup>g</sup></i> |
| 6                   | <b>10/11</b><br>1/11      | <b>I-M170</b><br><i>R-M207</i>       | <b>16-13-24-10-11-13</b><br><i>14-12-23-10-14-13<sup>g</sup></i> |
| 7                   | <b>14/16</b><br>2/16      | <b>I-M170</b><br><i>R-M207</i>       | <b>14-14-22-10-11-13</b><br><i>14-12-23-10-14-13<sup>g</sup></i> |

<sup>a</sup>Out of respect for the sensitivities of the present inhabitants of Tristan da Cunha, numbers were arbitrarily assigned to the families in order to avoid identification. <sup>b</sup>Number of males with the haplotype relative to the number of males tested in the family. <sup>c</sup>Haplogroups are defined by mutation as suggested by YCC.<sup>12</sup> <sup>d</sup>Microsatellite haplotypes are based on the alleles (repeat numbers) at the following loci in the order DYS19, DYS388, DYS390, DYS391, DYS392, DYS393. <sup>e</sup>Most likely derived from the haplotype associated with family 3 because of a single-step mutation at DYS391. <sup>f</sup>Due to introduction from the 'outside' – it does not match any Y lineages associated with surnames on the island. This haplotype differs from all other island haplotypes by the mutation SRY10831.2, a G to A reversion at the SRY10831 locus. <sup>g</sup>Most likely introduced from another family on the island.

*Note:* Haplotypes shown in bold are associated with the families or surnames on Tristan da Cunha. The two additional haplotypes are shown in italics.

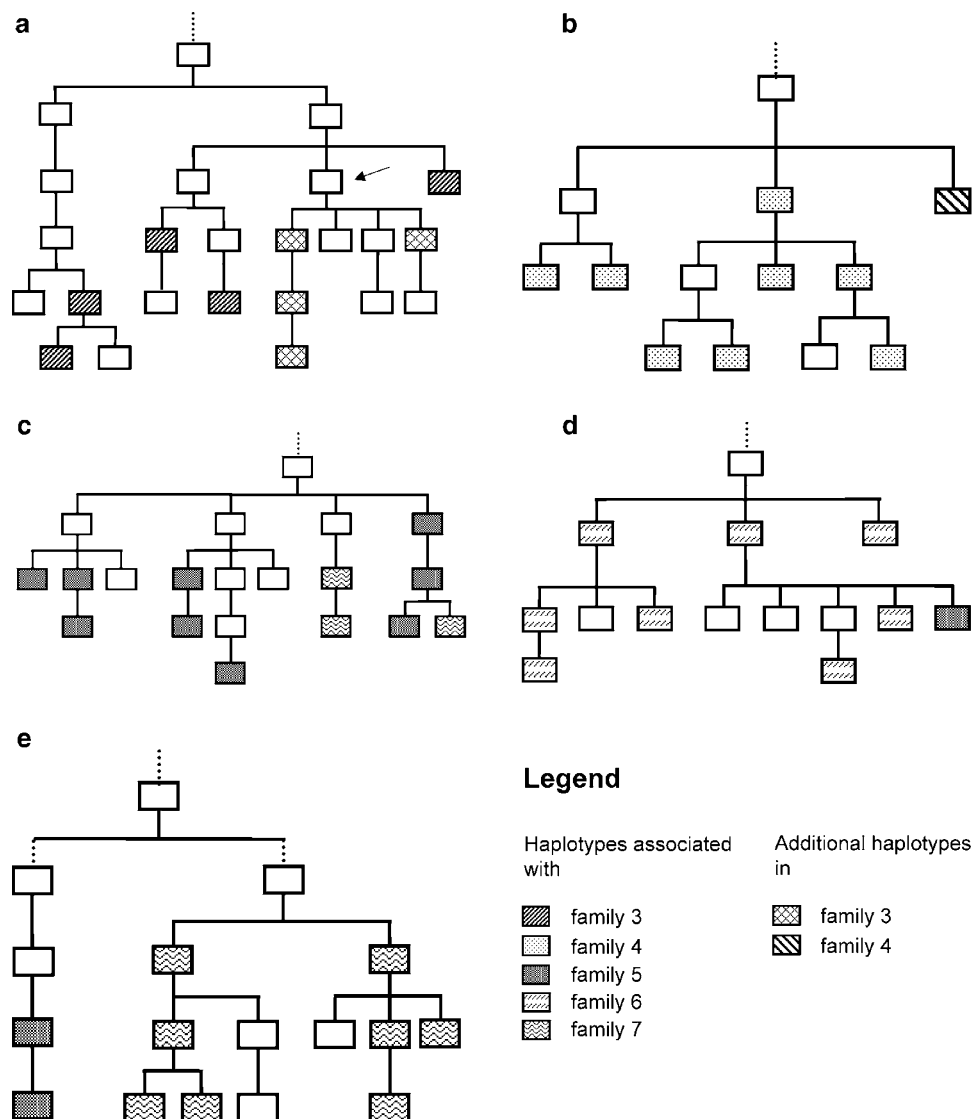
through transmission within family 3. Unfortunately, we do not have DNA from the parental generation where the mutation is presumed to have occurred, making paternity testing using autosomal markers impossible. Assuming the origin of the new haplotype within family 3, we estimated the mutation rate at the DYS391 locus to be 1 in 100 meioses ( $10 \times 10^{-3}$ , 95% CI  $0.2-54.5 \times 10^{-3}$ , excluding inconsistent transmissions, discussed later in the text). The mutation rate at this locus determined in a random sample of father/son pairs was reported to be  $4.82 \times 10^{-3}$  (95% CI  $0.61-16.4 \times 10^{-3}$ ).<sup>19</sup>

The second additional haplotype was observed in family 4, in one individual who had no sons (Figure 2b). This haplotype can be distinguished from all the others in the Tristan sample due to a reversion at SRY10831 on a haplogroup R-M207 background (Figure 1 and Table 1). This mutation defines a lineage that is particularly frequent in eastern and northern Europeans (54% in Polish, 47% in Russians, 31% in Norwegians).<sup>9</sup> Furthermore, the microsatellite allele constellation found in the Tristan individual (Table 1) represents the most common haplotype within

this haplogroup.<sup>20</sup> Based on genealogical information, we deduced that this new haplotype was introduced into family 4 in the early 1900s. It is well documented that many passenger ships, cargo vessels and whalers, some from Russia and Norway, used the island as a stopover port for trade and replenishing supplies.<sup>7</sup> Perhaps the new lineage was brought to the island during such a visit. This finding provides evidence for the contribution of a hidden ancestor who left his genes but not his name on the island.

The Y chromosome data in conjunction with genealogical information enabled us to unequivocally identify four additional instances of nonpaternity (in families 5–7; Table 1, haplotypes shown in regular font). In these cases, Y chromosomes were introduced that belonged to a different haplogroup from that associated with the family or surname. These 'newcomers', however, matched two other lineages already present on the island (Table 1 and Figures 2c–e). The 'new' chromosomes observed in two branches of the genealogy in family 5 are identical to the haplotype of family 7 (Figure 2c). The haplotype associated with family 5, on the other hand, was also found in families 6 (Figure 2d) and 7 (Figure 2e). There are two possible explanations for these findings: First, the introduced chromosomes could be the result of pre- or extramarital affairs involving men from the island community. There are indications in the records that would support this possibility.<sup>7</sup> The second explanation, of course, could be that the chromosomes were left on the island by other hidden ancestors who happened to have haplotypes that were identical to those of some of the islanders. In fact, the 'new' microsatellite haplotypes appear to be common in European populations; in the European Y-STR Haplotype Reference Database<sup>21</sup> the frequency for the haplotype of family 5 has been reported as  $1.5 \times 10^{-3}$  and for that of family 7 as  $3.9 \times 10^{-2}$  (the haplotype frequencies are calculated using five microsatellite loci only without considering DYS388 and the bi-allelic markers. As a result, the actual frequencies of the extended haplotypes used in this study would be lower). It is difficult to assess the probability of an internal versus an external contribution of these chromosomes to the Tristan population. However, given the presence of the two lineages in the community already (in families 5 and 7) and the geographic remoteness of Tristan da Cunha (thus the limited traffic to and from the island; Figure 3), it seems improbable that mysterious visitors contributed their chromosomes in all four instances.

We calculated the incidence of non-paternity in the male Tristan sample to be about 4% (five instances in 118 male births; illegitimate daughters cannot be detected studying Y chromosome transmission). Although this rate falls within the observed range of estimates (between 1.3 and 30%),<sup>3,22</sup> it is most likely underestimated, given the high level of inbreeding<sup>7</sup> and the consequently high extent of Y chromosome haplotype sharing on the island. Obviously,

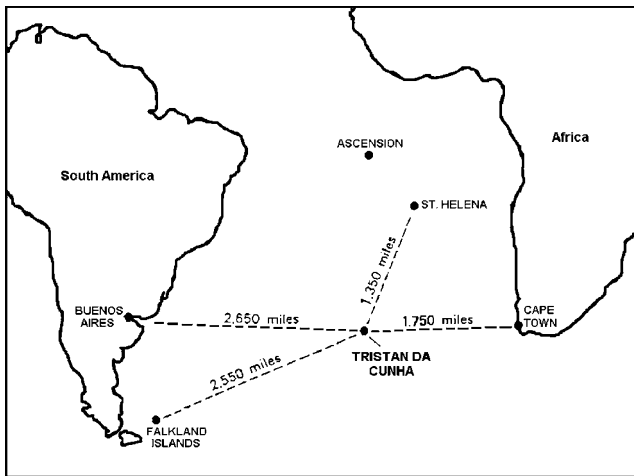


**Figure 2** Partial male genealogies showing inconsistencies in the transmission of Y chromosome haplotypes in five of the seven families on Tristan da Cunha. To conceal the identities of the families, only partial sib-ships are shown without indicating the generations in the pedigrees. Not all the males from whom Y chromosome data were obtained are presented. Tested individuals are indicated by filled squares. (a) Partial genealogy of family 3. An additional haplotype was found in four individuals spanning three generations. This haplotype most likely evolved from the haplotype associated with family 3 due to a single-step microsatellite mutation at DYS391. Arrow indicates the individual in whom the mutation probably originated. (b) Partial genealogy of family 4. The additional haplotype has a non-island origin (see Table 1). (c) Partial genealogy of family 5. The additional haplotypes observed in two branches of the genealogy are identical to the haplotype of family 7. (d) Partial genealogy of family 6. The additional haplotype in one branch of the genealogy is identical to the haplotype associated with family 5. (e) Partial genealogy of family 7. The additional haplotype is identical to the haplotype associated with family 5.

we could only discern those cases in which the Y chromosomes of the biological father and the name-giver were different. Interestingly, a study based on 27 autosomal serogenetic markers, including the highly informative Gm haplotype, did not indicate any inconsistencies in the genealogies of the Tristan da Cunha islanders.<sup>23</sup>

The uniqueness of the Tristan genealogy covering a period of almost 200 years presented us with an ideal

opportunity of testing the accuracy of written records with information stored in DNA. The Y chromosome data shown here, together with those based on mtDNA,<sup>8</sup> question the validity of some of the genealogical documentation. This study demonstrates convincingly that DNA is a powerful archival resource that can be used to test questions and to refine theories pertaining to human history and origins.



**Figure 3** Map showing the location of Tristan da Cunha in the South Atlantic and its relationship to neighbouring islands and continents. Tristan da Cunha is commonly referred to as the 'remotest island in the world'.

### Acknowledgements

We would like to thank Professor E du Toit for samples, Professor D Roberts for genealogical data as well as Ms C Dorf for technical assistance. This research was supported by grants from the National Health Laboratory Service (formerly known as South African Institute for Medical Research), the South African Medical Research Council, the National Research Foundation and the University of the Witwatersrand. AN is the recipient of the Hillel Friedland Postdoctoral Research Fellowship from the University of the Witwatersrand.

### References

- 1 Gill P, Ivanov PL, Kimpton C *et al*: Identification of the remains of the Romanov family by DNA analysis. *Nature Genet* 1994; **6**: 130–135.
- 2 Foster EA, Jobling MA, Taylor PG *et al*: Jefferson fathered slave's last child. *Nature* 1998; **396**: 27–28.
- 3 Sykes B, Irven C: Surnames and the Y chromosome. *Am J Hum Genet* 2000; **66**: 1417–1419.
- 4 Jobling MA: In the name of the father: surnames and genetics. *Trends Genet* 2001; **17**: 353–357.
- 5 Roberts DF: Genetic effects of population size reduction. *Nature* 1968; **220**: 1084–1088.

- 6 Department for Environment, Food and Rural Affairs (DEFRA), UK Overseas Territories, consensus data for 2001, <http://www.defra.gov.uk/science/geneticresources/access/overseas.htm>.
- 7 Roberts DF: The development of inbreeding in an island population; in Morris LN (ed): *Human populations, genetic variation and evolution*. San Francisco, Chandler Publishing Company, 1971, pp 381–391.
- 8 Soodyall H, Jenkins T, Mukherjee A, du Toit E, Roberts DF, Stoneking M: The founding mitochondrial DNA lineages of Tristan da Cunha islanders. *Am J Phys Anthropol* 1997; **104**: 157–166.
- 9 Rosser ZH, Zerjal T, Hurles ME *et al*: Y-chromosomal diversity in Europe is clinal and influenced by geography, rather than by language. *Am J Hum Genet* 2000; **67**: 1526–1543.
- 10 Semino O, Passarino G, Oefner PJ *et al*: The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 2000; **290**: 1155–1159.
- 11 Underhill PA, Shen P, Jin L *et al*: Y chromosome sequence variation and the history of human populations. *Nature Genet* 2000; **26**: 358–361.
- 12 YCC – the Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; **12**: 339–348.
- 13 Hammer MF, Horai S: Y chromosomal variation and the peopling of Japan. *Am J Hum Genet* 1995; **56**: 951–962.
- 14 Santos FR, Pandaya A, Tyler-Smith C *et al*: The central Siberian origin for native American Y chromosomes. *Am J Hum Genet* 1999; **64**: 619–628.
- 15 Underhill PA, Jin L, Lin AA *et al*: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; **7**: 996–1005.
- 16 Shen P, Wang F, Underhill P *et al*: Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci USA* 2000; **97**: 7354–7359.
- 17 Jobling MA, Tyler-Smith C: Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 1995; **11**: 449–456.
- 18 Kayser M, Caglia A, Corach D *et al*: Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 1997; **110**: 125–133.
- 19 Kayser M, Roewer L, Hedman M *et al*: Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 2000; **66**: 1580–1588.
- 20 Zerjal T, Pandya A, Santos FR *et al*: The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. in Papiha SS, Deka R, Chakraborty R (eds): *Genomic diversity: applications in human population genetics*. New York, Plenum Press, 1999, pp 91–101.
- 21 Y-STR Haplotype Reference Database, <http://ystr.charite.de>
- 22 MacIntyre S, Sooman A: Non-paternity and prenatal genetic screening. *Lancet* 1991; **338**: 869–871.
- 23 Jenkins T, Beighton P, Steinberg AG: Serogenetic studies on the inhabitants of Tristan da Cunha. *Ann Hum Genet* 1985; **12**: 363–371.