Human Molecular Genetics, 1996, Vol. 5, No. 11 **1767–1775**

Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males

Mark A. Jobling*, Vassiliki Samara, Arpita Pandya[1], Neale Fretwell, Barbara Bernasconi[1,+], R. John Mitchell[2], Tudevdagva Gerelsaikhan[3], Bumbein Dashnyam[3], Antti Sajantila[4], Pia J. Salo[5], Yutaka Nakahori[6], Christine M. Disteche[7], Kumarasamy Thangaraj[8], Lalji Singh[8], Michael H. Crawford[9] and Chris Tyler-Smith[1]

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK, [1]CRC Chromosome Molecular Biology Group, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK, [2]Department of Genetics and Human Variation, La Trobe University, Bundoora Campus, Bundoora, Victoria 3083, Australia, [3]Department of Molecular Biology, Institute of Biotechnology, Academy of Sciences, Ulaanbaatar-51, Mongolia, [4]Department of Forensic Medicine, University of Helsinki, PO Box 40, FIN 00014 Helsinki, Finland, [5]Department of Medical Genetics, University of Helsinki, PO Box 21, FIN 00014 Helsinki, Finland, [6]Department of Human Genetics, School of International Health, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan, [7]Department of Pathology, University of Washington, Seattle, WA 98195, USA, [8]CCMB, Uppal Road, Hyderabad 500 007, India and [9]Department of Anthropology, University of Kansas, Lawrence, KA, USA

Received June 11, 1996; Revised and Accepted August 28, 1996

Deletion of the 50f2/C (*DYS7C*) locus in interval 6 of Yq has previously been reported as a polymorphism in three males. We describe a survey of worldwide populations for further instances of this deletion. Of 859 males tested, 55 (~6%) show absence of the 50f2/C locus; duplication of the locus was also detected in eight out of 595 males (~1.4%). Populations having the deletion are confined to Asia, Australasia, and southern and northern Europe; of those of reasonable sample size, Finns had the highest deletion frequency (55%; *n* = 21). The deletions vary in size and the larger ones remove some of the *RBM* (RNA Binding Motif) genes, but none of the deletion males lack *DAZ* (Deleted in AZoospermia), a candidate gene for the azoospermia factor. On a tree of Y haplotypes, 28 deletion and eight duplication chromosomes fall into six and four haplotypic groups respectively, each of which is likely to represent an independent deletion or duplication event. Microsatellite and other haplotyping data suggest the existence of at least two further classes of deletion. Thus duplications and deletions in this region of Yq have occurred many times in human evolution, but remain useful markers for paternal lineages.

INTRODUCTION

Because it is uniparentally inherited and escapes from recombination along most of its length, the Y chromosome is a useful tool for the study of human paternal lineages (1). Analysis of DNA polymorphisms allows the construction of haplotypes which can be used to build trees of Y chromosomes, and comparisons of the frequencies of these haplotypes in different populations can give information about population histories and relationships. Although several new polymorphisms have been reported recently (2–5), there is still a need for many more.

Three insertion/deletion type polymorphisms have been reported on the Y chromosome. Two of these, the insertion of an *Alu* element [*DYS287* (6)] since the human–chimp divergence, and a 2 kb insertion or deletion detected by the probe 12f2 [*DYS11* (7)], appear to represent unique events in human evolution, and are useful in the construction of trees and in population analysis.

Deletion of the 50f2/C (*DYS7C*) locus was previously reported as a polymorphism in a single male of Finnish origin, who was the normal father of an XY female (8). A subsequent survey of 200 normal Japanese males revealed two more cases (9) but there has been no evaluation of the usefulness of this marker for evolutionary purposes. We have now searched in worldwide samples for further instances of such deletions, and addressed the question of whether all males having a deletion of 50f2/C share it by descent.

As well as its potential as an evolutionary tool, deletion of 50f2/C is relevant to the issue of male infertility: genetic evidence has existed for many years for the presence in the distal euchromatin of Yq of a gene or genes (azoospermia factor—*AZF*) essential for spermatogenesis (10), and *de novo* interstitial deletions, some of which include 50f2/C (11), are associated with azoospermia. We ask whether any of the candidate genes for *AZF* are removed with 50f2/C in normal males, and our study also contributes to knowledge of the normal variation in the structure

*To whom correspondence should be addressed
+Present address: Via Majella 25, 21100 Varese, Italy

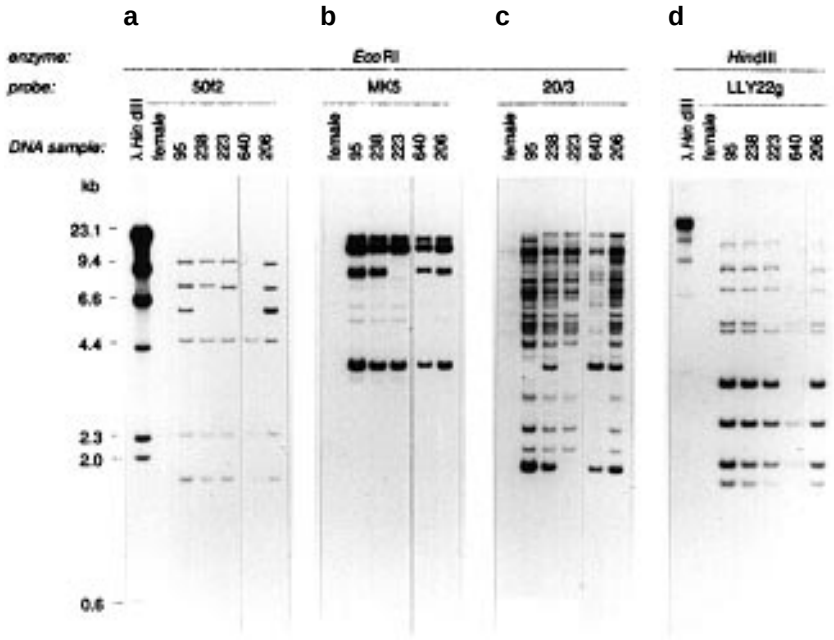

Figure 1. Autoradiographs showing the results of hybridisations of four probes to digests of DNAs from a control female, a control male (m95), the three deletion males m238, m223 and m640, and the duplication male m206. (**a**) Probe 50f2, *Eco*RI digests: the 50f2/C fragment is absent in m238, m223, and m640, and is present at increased intensity in m206. The 50f2/A and B fragments are also absent in m640. (**b**) Probe MK5 (*RBM* cDNA), *Eco*RI digests: a 9.5 kb band is absent from m223 alone, defining the 'large' deletion. (**c**) Probe 20/3, *Eco*RI digests: a 1.9 kb band is absent from the same male. Presence or absence of the strong 4.1 kb band is an unrelated polymorphism; 20/3 detects multiple fragments on Yp, as well as Yq, and absence of some of these from m640 is likely to be the result of the short arm deletion in this male. (**d**) Probe LLY22g, *Hin*dIII digests: the D fragment (3.5 kb) is absent from the 'large' deletion male m223 only. Absence of the A, B, C and F fragments from m640 is a result of his short arm deletion, and the faintness of other signals in this sample is due to underloading. Markers are*Hin*dIII fragments of phage λ DNA.

of Yq, which is an essential basis for interpreting the significance of rearrangements which apparently have phenotypic consequences.

RESULTS

Survey of worldwide samples for 50f2/C deletions

We used two methods, based on hybridisation and PCR, to search for 50f2/C deletions in 859 males from 46 different populations. The probe 50f2 (*DYS7*) detects five Y-chromosomal loci, seen as five *Eco*RI fragments in hybridisation analysis (Fig. 1a); 50f2/C is the third largest of these, and maps to interval 6E on Yq (12). 595 males were tested using this hybridisation method, and 32 new deletions detected. For the remainder (264 males), sufficient DNA was not available for hybridisation, and so a PCR assay was used, detecting a further 23 new deletions. In all cases where sample availability permitted, the two assays were shown to be concordant: all (35/35) deletion males defined by hybridisation were shown to be deleted by the criterion of the PCR assay, and 50/50 males positive by hybridisation were also positive by PCR. In these, and in subsequent experiments, the three previously described deletion males (m121, ref. 8; cases 920 and 1423, ref. 9) were also included.

The 55 new deletion males (Table 1) belong to 12 different populations from Asia, Europe and Oceania. Observed population frequencies of deletions vary from 0 to >50%. The first deletion male to be described (8) was of Finnish origin. For this reason Finns were included in our survey, and were found to have the highest frequency of deletions of any population of reasonable sample size: 11 of 20 males had a deletion. In hybridisation analysis, one Sri Lankan male (m640) showed deletion of the 50f2/A and B loci, mapping to Yp (13), as well as 50f2/C. This could be due either to two separate deletion events, or to a pericentric inversion bringing the three loci close together, followed by a single deletion.

Duplication of the 50f2/C locus

As well as deletions of 50f2/C, hybridisation analysis of 595 males revealed eight cases where the intensity of the 50f2/C *Eco*RI fragment was greater than normal (Table 1). This was confirmed by densitometry, and shown to be consistent with duplication of 50f2/C. The PCR assay did not demonstrate these duplications convincingly. Duplication males are found in the Mongolian and Indian populations, and also in an Algerian, a San, and in two Adygeans (from south-western Russia).

Markers co-deleted with 50f2/C

Are the 58 deletions the result of a single event in human evolution? Two approaches were taken to address this issue. In the first, loci were sought which map to the same region of the chromosome as 50f2/C and are deleted in some, but not all, deletion males. This would reflect differences in the sizes and, therefore, probably the origins of the deletion events.

In hybridisation analysis, three probes, 20/3, LLY22g, and MK5 [cDNA for the gene *RBM* (RNA Binding Motif), previously known as *YRRM* (Y chromosome RNA Recognition Motif); 14] detect major bands which are absent from a sub-set of the deletion males found only in Mongolia and China (m223 in Fig. 1b, c and d). Minor bands detected by the probes 48, 94/3, 27, 36b and RBF2 were also absent (data not shown). Two classes of deletions are thus

Table 1. Results of population screening for 50f2/C deletions and duplications

Continent	Population	No. tested	No. deleted	No. tested	No. duplicated
Europe	UK	39	0	39	0
	Italian	32	0	12	0
	French	4	0	3	0
	German	2	0	2	0
	Portuguese	1	0	1	0
	Basque	30	0	30	0
	Finnish	20	11	0	0
	Norwegian	10	1	10	0
	Swedish	1	0	1	0
	Saami	9	3	0	0
	Greek	22	2	0	0
	Ukrainian	1	0	0	0
	Bulgarian	5	0	5	0
	other	15	0	15	0
Asia	Chinese	53	5	52	0
	Japanese	15	0	2	0
	Malaysian	16	0	1	0
	Cambodian	1	0	1	0
	Indian	220	1	215	3
	Nepalese	1	0	1	0
	Tibetan	3	0	3	0
	Pakistani	2	0	2	0
	Sri Lankan	25	1	25	0
	Mongolian	65	12	65	1
	Indonesian	17	0	0	0
	Altai	28	5	0	0
	Yakut	5	5	5	0
	Kajikistani	4	0	0	0
	Adygean	4	0	4	2
	other	2	0	2	0
Africa	Kenyan	14	0	14	0
	Nigerian	3	0	0	0
	Zimbabwean	34	0	0	0
	Ghanaian	1	0	1	0
	San	9	0	9	1
	Biaka Pygmy	4	0	4	0
	Mbuti Pygmy	1	0	1	0
	Berber	1	0	0	0
	Algerian	49	0	49	1
	other	21	0	4	0
Oceania	Australian	46	8	5	0
	Cook Islander	19	0	0	0
	Melanesian	2	0	2	0
	Papua New Guinean	11	1	0	0
America	Karitiana	1	0	1	0
	Surui	18	0	3	0
	Mayan	1	0	1	0
	Navaho	1	0	1	0
	Tohono O'Odham	1	0	1	0
	other	2	0	2	0
Total		859	55	595	8

Deletion testing was done by PCR or hybridisation analysis or both, and duplication testing by hybridisation analysis only. The previously described deletion cases are not included.

defined, which are referred to as 'large' and 'small'; four of 28 males tested had 'large' deletions. Male 640, deleted for 50f2/A and B as well as C, shows normal hybridisation patterns for MK5 (Fig. 1b); in the case of LLY22g (Fig. 1d), the band (D) deleted in the 'large' class of males is present, but four others (A, B, C and F) are absent. These map to Yp (13), and are likely to have been removed together with 50f2/A and B. Some bands detected by 20/3 are also absent from this male (Fig. 1c), and the same explanation applies.

Fourteen established and three novel Y-specific STSs (sequence-tagged sites) from the 50f2/C region were also screened in PCR assays for co-deletion with 50f2/C. Because it was of particular interest to see whether *DAZ* (Deleted in AZoospermia), a candidate gene for *AZF*, was absent from any of the deletion males, two STSs from the *DAZ* gene itself (15) were included. None of the STSs tested were deleted from any of the males, and this analysis is therefore uninformative, but the hybridisation analysis shows that the deletions are heterogeneous: the 'large' and 'small' deletions are likely to represent at least two events.

Markers co-duplicated with 50f2/C

One of the individuals carrying a duplication of 50f2/C, the San male m720, also showed an increase in intensity of the 50f2/E fragment consistent with a duplication of this locus. 50f2/E lies on the long arm of the Y chromosome, proximal to 50f2/C in interval 6A (16), and is probably co-duplicated in a single event with 50f2/C. Hybridisation using the probe LLY22g shows a duplication of the E fragment, which, from analysis of 50f2-containing YAC clones, lies close to 50f2/E (data not shown).

Compound haplotyping of deletion and duplication chromosomes

The second approach taken to the issue of heterogeneity of origin of the deletions, and also the duplications, was to use Y-chromosomal haplotyping. Combinations of polymorphisms which are likely to represent very rare or unique events define haplotypes which can be used to build trees of Y chromosomes (1). We wanted to use haplotyping to ask where on such a tree the deletion and duplication chromosomes lay. Because many of these polymorphisms cannot be assayed by PCR, and rely on traditional methods, this approach was feasible only for samples where large amounts of high molecular weight DNA were available. Twenty-eight of the 58 deletion chromosomes, and all eight of the duplication chromosomes were analysed.

Figure 2 shows a tree of Y chromosome haplotypic groups constructed using 15 polymorphisms, and the positions upon it of 50f2/C deletion and duplication cases; the group to which each individual belongs is shown in Table 2.

The 'large' deletions occur in two different haplotypic groups, and the 'small' deletions in four. Thus both the large and small deletions are themselves heterogeneous. A single haplotype, group 12, accounts for 19 of the 28 deletion chromosomes. Deletions associated with the remaining five groups are rare, and four (m38, m204, m238 and m640) occur as singletons. The eight duplication chromosomes belong to four different haplotypic groups.

Diversity within deletion and duplication classes defined using microsatellites

Four of the six different haplotypic groups associated with deletions of 50f2/C are found in a single population, the Mongolians. Do other populations display similar levels of heterogeneity, and can we find evidence of further independent deletion events? Since amounts of DNA were limiting, haplotyping with Y-specific microsatellites, which can be assayed by PCR, was used to address this question. Because of their relatively high mutation rates, microsatellites also give a measure of the diversity within the already defined haplotypic groups.

Using one four-locus (*DYF371*) and four single-locus microsatellites (*DYS19, DYS390, DYS425* and *DYS426*), we first determined haplotypes for the 28 deletion chromosomes which were already well characterised, and seven of the eight duplication chromosomes. These data are summarised in Table 2.

The 28 deletion chromosomes have eight different microsatellite haplotypes. Correlation between these and the six groups already defined is good. The three singletons m38, m204 and m640 each have unique haplotypes, as does m238, emphasising the distinctness of this chromosome from those in the two Australian males which belong, with m238, to group 2 within the tree. The Australians have microsatellite haplotypes not seen elsewhere, which differ from each other by a single repeat unit change at *DYS19*. The three group 10L Mongolians with the 'large' deletion share a haplotype not seen in 212 non-deletion males tested. Seventeen of the 19 common group 12 deletion chromosomes share a single microsatellite haplotype, which has not been observed elsewhere, and the remaining two have a haplotype which differs from this only by a single repeat unit gain at *DYS390*.

The microsatellite data on the seven typed duplication chromosomes divide the four group 2 chromosomes into three distinct haplotypic classes, suggesting that there may be more than one duplication event represented in this group. However, we cannot exclude the possibility that these chromosomes share their duplications by descent.

Microsatellite haplotypes were next determined for the remaining 30 deletion chromosomes. The six newly analysed Australians share a haplotype which stands out as being different from those of all other deletion chromosomes: they may therefore possess a distinct deletion. Of the remainder, 13 chromosomes have one of the two haplotypes previously shown to be associated with group 12, and five have a haplotype differing from these only by a single repeat unit gain or loss at *DYS19* or *DYS390*. One of the Greeks has the haplotype seen in the Mongolian (group 10L) 'large' deletions, and three of the Altai, one of the Japanese and the Papua New Guinean have haplotypes which differ from this only by one or two repeat unit gains at *DYS19* and single repeat unit gains at *DYS390*. In addition, the prediction that these haplotypes are associated with a 'large' deletion could be tested in the case of the Greek and Altai samples by hybridisation using LLY22g (data not shown), and was borne out—the D fragment was absent in each case. The six Australians could also be tested, and showed a normal LLY22g hybridisation pattern, indicating that their deletion is 'small'.

The finding of identical microsatellite haplotypes shared by deletion chromosomes within populations such as the Australians (6/8), the Yakut (5/5) and the Finns (8/11) suggests that these males may also share recent patrilineal descent. This possible

Table 2. Haplotypes of deletion and duplication chromosomes

Male	Population	Microsatellites					Haplotypic group	
		DYS426	*DYF371*	*DYS425*	*DYS19*	*DYS390*	observed	deduced
Deletion chromosomes								
m119	Australian	94	204, 195	204	190	199	2	
m120	Australian	94	204, 195	204	194	199	2	
m203	Mongolian	94	201, 198	201	190	211	12	
m204	Mongolian	97	210, 201, 195	201	186	211	1	
m205	Mongolian	94	201, 198	201	190	211	12	
m209	Mongolian	94	201, 198	201	190	211	12	
m223	Mongolian	94	204, 201, 195	201	194	215	10L	
m227	Mongolian	94	201, 198	201	190	211	12	
m229	Mongolian	94	204, 201, 195	201	194	215	10L	
m238	Mongolian	94	210, 198	198	190	215	2	
m243	Mongolian	94	201, 198	201	190	211	12	
m252	Mongolian	94	204, 201, 195	201	194	215	10L	
m253	Mongolian	94	201, 198	201	190	211	12	
m274	Mongolian	94	201, 198	201	190	211	12	
m462	Indian	94	201, 198	201	190	215	12	
m640	Sri Lankan	94	210, 201, 195	201	194	207	10AB	
m38	Chinese	94	210, 204, 201	201	194	211	2L	
m297	Chinese	94	201, 198	201	190	211	12	
m300	Chinese	94	201, 198	201	190	211	12	
m307	Chinese	94	201, 198	201	190	211	12	
m326	Chinese	94	201, 198	201	190	211	12	
m121	Finnish	94	201, 198	201	190	215	12	
m295	Norwegian	94	201, 198	201	190	211	12	
m721	Yakut	94	201, 198	201	190	211	12	
m722	Yakut	94	201, 198	201	190	211	12	
m723	Yakut	94	201, 198	201	190	211	12	
m724	Yakut	94	201, 198	201	190	211	12	
m725	Yakut	94	201, 198	201	190	211	12	
LGL5144	Finnish	94	201, 198	201	190	215		12
LGL5176	Finnish	94	201, 198	201	186	211		12
LGL5190	Finnish	94	201, 198	201	190	215		12
LGL5191	Finnish	94	201, 198	201	194	211		12
LGL5198	Finnish	94	201, 198	201	190	215		12
LGL5209	Finnish	94	201, 198	201	190	215		12
LGL5236	Finnish	94	201, 198	201	190	215		12
LGL5246	Finnish	94	201, 198	201	190	215		12
LGL5254	Finnish	94	201, 198	201	190	215		12
LGL5293	Finnish	94	201, 198	201	190	219		12
LGL5298	Finnish	94	201, 198	201	190	215		12
S10	Saami	94	201, 198	201	190	211		12
S11	Saami	94	201, 198	201	186	215		12
S15	Saami	94	201, 198	201	190	215		12
Au2	Australian	94	207, 204, 201, 195	201	202	211		?
Au4	Australian	94	207, 204, 201, 195	201	202	211		?
Au7	Australian	94	207, 204, 201, 195	201	202	211		?
Au9	Australian	94	207, 204, 201, 195	201	202	211		?
Au11	Australian	94	207, 204, 201, 195	201	202	211		?
Au18	Australian	94	207, 204, 201, 195	201	202	211		?
G894	Greek	94	204, 201, 195	201	194	215		10L
G896	Greek	94	201, 198	201	190	211		12
A19	Altai	94	201, 198	201	190	215		12
A33	Altai	94	204, 201, 195	201	198	215		10L
A72	Altai	94	201, 198	201	194	211		12
A86	Altai	94	204, 201, 195	201	198	215		10L
A106	Altai	94	204, 201, 195	201	198	219		10L
920	Japanese	94	204, 201, 195	201	202	219		10L
1423	Japanese	94	201, 198	201	190	211		12
GIMI18	Papua New Guinean	94	204, 201, 195	201	194	219		10L
Duplication chromosomes								
m206	Mongolian	94	204, 201, 195	201	190	215	2	
m388	Indian	94	210, 204, 195	204	190	207	2	
m405	Indian	97	210, 207, 204, 201, 195	201	n.d.	211	1	
m416	Indian	97	210, 207, 204, 201, 195	201	n.d.	207	1	
m670	Algerian	n.d.	n.d.	n.d.	n.d.	n.d.	9	
m720	San	97	204, 201, 195	null	202	191	7	
m727	Adygean	94	207, 204, 195	207	194	211	2	
m729	Adygean	94	207, 204, 195	207	194	211	2	

Allele sizes (in bp) are shown for the microsatellites tested. For *DYF371*, there are four (rarely, five) loci on the Y chromosome; often, fewer than four products are visible, due to comigration or absence of one or more products. The *DYS425* assay amplifies one of the *DYF371* loci specifically: the product size for this assay is actually 94 bp smaller than that of the same locus amplified with the *DYF371* four-locus assay, but, to avoid confusion, the size given in the table is the corresponding *DYF371* size. The haplotypic groups observed directly for 28 deletion and all eight duplication chromosomes are shown (see Fig. 2), as are those deduced from the microsatellite haplotypes of the remaining 30 deletion chromosomes. The suffix 'L' denotes a 'large' deletion, and the suffix 'AB' denotes deletion of 50f2/A and B loci. '?' indicates that the haplotypic group is unknown. 'null' (microsatellite *DYS425*) indicates that a PCR assay failed to give a product in at least three attempts, and may be due to deletion of the locus or mutation in a primer site.
n.d., not done.

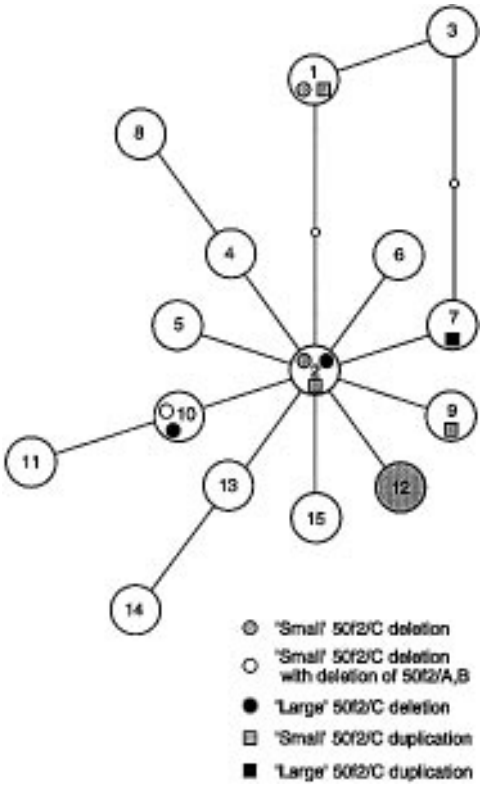

Figure 2. Unrooted parsimony tree of 15 Y chromosome haplotypic groups, showing the positions of deletion and duplication chromosomes. Large circles represent haplotypic groups, and single lines between these represent single mutation events. The 'large' deletions occur in two groups, the 'small' in four groups, and the duplications in four groups. All of the group 12 chromosomes have a small deletion of 50f2/C, and this is indicated by shading of the large circle in this case. Fourteen polymorphisms which are thought to represent unique events in human evolution were used to construct this tree, which is based on one published previously (1). Haplotypic groups are numbered in their order of discovery, and some used in the original tree have now been renamed: group 5 was previously 'J', 6 = 'P', 7 = '!K', 8 = 'Af', 9 = 'Med'. Because of ambiguity about the status of one polymorphism, *SRY* (ref. 1; unpublished data), part of the tree is shown as a network. Novel polymorphisms used in tree construction will be described elsewhere (A.Pandya *et al.*, manuscript in preparation).

problem of sampling means that regional figures for deletion frequencies should be regarded with caution. However, in an independent sample of Finns, seven of 18 males were found to be deleted (P.J.Salo, unpublished data), and so in this case, at least, the high frequency we have observed may well be typical of the population as a whole.

DISCUSSION

We have found 55 apparently normal males who lack the 50f2/C locus in the distal euchromatin of Yq. In order to understand the origins of these chromosomes and the three previously known 50f2/C deletions, we constructed haplotypes using polymorphisms representing unique or rare mutations (28 chromosomes) and using microsatellites (all 58 chromosomes). The resulting haplotypes (Table 2; Fig. 2) show that some polymorphisms must have arisen more than once. We can use information on the mutation rates of the different types of polymorphism to determine the likely number of origins of the 50f2/C deletion. Base substitutions have a very low mutation rate [~10^{-9} per base per year, or 3×10^{-4} per base per generation (3,4)], and microsatellites a high rate [~10^{-3} per locus per generation (17)]. The rate expected for a Yq deletion can be estimated from the frequency of deletions leading to infertility. This is 10^{-4} per generation (18), but since at least three distinct loci are involved, an average deletion rate for a Yq locus would be 3×10^{-5} per generation. Although these figures are imprecise, it is considerably (~1000-fold) more likely that a deletion will have recurred than a point mutation. Thus, when deletions occur as subsets of the chromosomes in Y groups (Fig. 2) they are likely to represent independent events. Taking into account the 'large' and 'small' sizes of the deletions, such reasoning would suggest a minimum of five deletion events. This minimum number would require that the group 12 chromosomes (which are all deleted for 50f2/C), two of the Australian deletion chromosomes (group 2 in the tree) and m238 (also group 2), share their deletions by descent. The group 2 Australian chromosomes differ from m238 at three and four of the five microsatellite loci tested and also at other loci (results not shown); taking into account their distinct geographical location, this suggests a separate deletion event. It is more difficult to decide whether the group 12 chromosomes could share the m238 50f2/C deletion. For some chromosomes the microsatellites differ at only two of the five loci, and the geographical ranges overlap. These considerations therefore suggest the existence of either six or seven deletion events. The final set of distinct deletion chromosomes is represented by the Australians Au2–Au18. They differ from all the other deletion chromosomes at the multilocus microsatellite *DYF371*, but are otherwise most similar to the Japanese (920) group 10L chromosome. However, the finding that they do not possess a 'large' deletion makes it seem likely that they represent an independent 50f2/C deletion event. In conclusion, our data suggest that there have been seven or eight 50f2/C deletions.

It is easier to evaluate the number of duplication events because there are fewer chromosomes and haplotyping information is available for all of them. They form subsets of four different haplotypic groups, implying the existence of at least four distinct duplication events.

Is the large number of 50f2/C duplication and deletion events in the population (~7%) typical of other parts of the Y chromosome? Triplications of *DYS19* (19), and duplication and triplication of the pDP31 (*DXYS1Y*) locus (20) exist, and we have found a deletion of 50f2/A and B. The proportion of chromosomes carrying deletions and duplications must be high, and the action of these dynamic processes over long periods of time helps to explain why there are so few unique sequences on the Y.

Little can be said from these studies about the sizes of the deleted or duplicated regions: the duplication in m720 includes the 50f2/E locus, and is therefore likely to be on the megabase scale, and preliminary studies of YAC clones containing 50f2/C as well as the loci which are co-deleted in the 'large' deletions suggests that these events involve at least several hundreds of kilobases. The mechanisms at work may be similar to those of deletions which lead to azoospermia through the loss of genes in Yq, and are likely to involve aberrant recombination events between repeat sequences, in which the region is particularly rich.

Positional cloning of the *AZF* locus (or loci) has proved difficult. Two candidate genes have been identified: *RBM* (21) is a gene family with about 30 very similar members on Yp as well as Yq, some of which are absent from infertile males, but it has been difficult to determine which, if any, of these are essential for spermatogenesis. *DAZ* is reported to be a single-copy locus (15); however, although 12 *de novo* deletions including this gene were

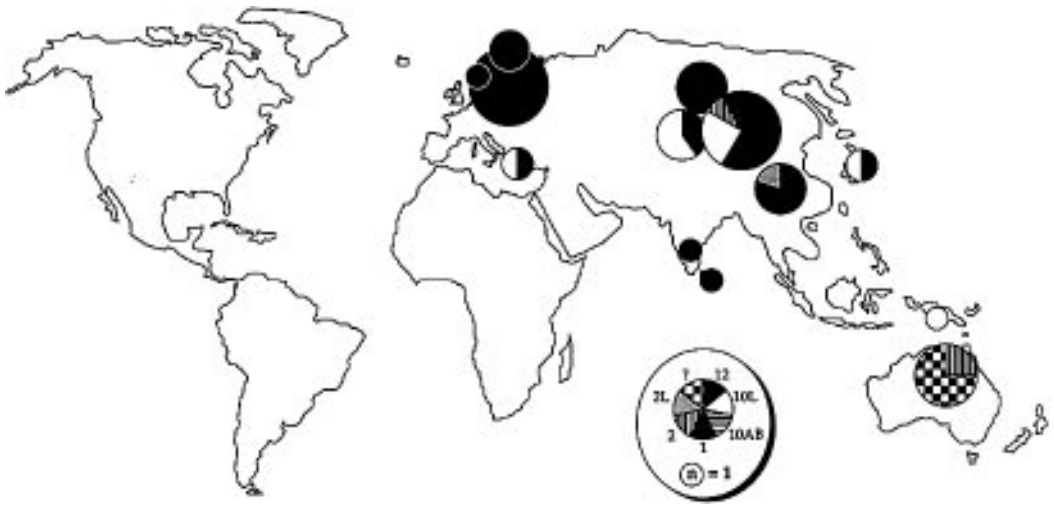

Figure 3. World distribution of different classes of 50f2/C deletions. The classes of deletion within each of 13 populations are indicated by the pie charts. The area of each chart is proportional to the number of individuals. Seven classes are here defined by haplotypic group; however, the deletions in haplotypic group 2 found in Mongolians and Australians are likely to have different origins (see text).

found in infertile males, point mutations have yet to be described. Our results illustrate the importance of demonstrating that deletions are *de novo* events. The absence of some of the *RBM* sequences from the 'large' classes of deletion indicates that these members of this gene family, at least, are dispensable. None of the deleted males lack *DAZ*, which is consistent with this gene being *AZF*.

Despite the recurrence of 50f2/C deletions, they can be combined with other markers to provide important insights into population history. Thirty-seven of the total of 58 deletion males belong to a single haplotypic group, group 12, and therefore represent a relatively common and geographically widespread class of deletion chromosome, stretching from Siberia in the north-east of Asia, south to China, Japan, and India, and west to Scandinavia (Fig. 3). The high frequency (55%) in Finns is particularly striking, and appears to show an affinity between Finnish, Siberian and Mongolian Y-chromosomal lineages; this is in contrast to the data from autosomal and mitochondrial DNA studies (22,23), which show Finns to be closely related to Indo-European speakers such as Italians and Germans, and is more consistent with traditional linguistic studies which suggest that the Finns have their origins in a 'Finno–Ugric homeland', either in the Altai, Uralic or Volga regions (24). The history of Y chromosomes is distinct from the histories of other parts of the genome, and is influenced by the phenotype that the Y confers, i.e., maleness. Thus, these observations may reflect historical movements, predominantly of males, across Northern Asia and Europe.

Alternatively, the high frequency of these deletion chromosomes in the Finns may be the result of drift in a population which had been founded by only a few individuals, or, as some Y-chromosomal and mitochondrial DNA data suggest (25), subject to a bottleneck in its early history. Similar explanations have been proposed for the high incidences of some genetic disorders in the Finns (26). The finding of these group 12 chromosomes in three of nine Saamis is more in agreement with findings from other markers, which show Saamis as outliers in Europe (22,23). A second and relatively common deletion class (group 10L) exists, in nine of the 58 males, with a more easterly distribution (Fig. 3). The rare deletions and the duplications (which are difficult to score) appear less informative, but may prove to be useful markers for populations we have not sampled here. There is no reason to suppose that we have discovered all of the existing classes of deletion; indeed, the representation of several classes by single chromosomes in our collection suggests that other classes remain undetected. The inclusion of the 50f2/C PCR assay in population studies in the future may reveal more.

Markers representing unique mutational events have been shown to be useful for Y evolutionary studies (3), but few are available. In addition to loci like 50f2/C, there are now more than 20 microsatellites known on the Y chromosome (1; this study), and used in combination they can generate very many haplotypes. Although the construction of trees using such haplotypes alone is problematic (27), we have demonstrated the value of using them within the framework of a robust tree defined using unique or rare event polymorphisms. The availability of many different classes of polymorphism with different mutation rates is one of the features which makes the Y chromosome such a useful tool for studying human evolution.

MATERIALS AND METHODS

Standard methods were as described previously (28,29).

Genomic DNA samples

The 859 samples screened included those previously described (30,31), several collections made by the authors, Basque and Algerian DNAs kindly donated by Adolfo López de Munain and Doudja Nafa respectively, several samples kindly donated by Alec Jeffreys, and a sub-set from the Y Chromosome Consortium Repository (32). Male 121 is the first deletion case to be described,

the father of patient 2 in ref. 8, and samples 920 and 1423 are the two previously discovered Japanese deletion males (9).

Probes

Sources, references, and conditions for the probes 50f2 (*DYS7*), 20/3 (*DYS26*), 94/3 (*DYS35*), 48 (*DYS6*), 36b (*DYS4*), 27 (*DYS5*) and RBF2 can be found in ref. 30. Details of LLY22g are given in ref. 11. MK5 (the *RBM* cDNA) was a gift from Ma Kun (MRC Human Genetics Unit, Edinburgh), and was used as described (14).

Haplotyping

Eight of the 14 polymorphisms used to construct the haplotypic tree (Fig. 2), and methods for their typing, have been described previously (1); six are novel, and will be described elsewhere (A.Pandya *et al*., manuscript in preparation). Further details of these polymorphisms are available from C.Tyler-Smith on request.

Densitometric analysis

Filters showing putative duplications of 50f2/C were scanned using the Packard InstantImager or the Molecular Dynamics PhosphorImager, and bands quantitated using the manufacturers' supplied software.

PCR assays

All reactions were in a 10 µl volume and used 1.0 U *Taq* polymerase (Advanced Biotech) and 50–100 ng template DNA in the buffer described (33).

50f2/C deletion. PCR primers (1) designed to amplify the Y-specific minisatellite *MSY1* [*DYF155S1* (34)], which is adjacent to the 50f2/B locus, coamplify a 196 bp product (*DYF155S2*) lying 4 kb from the 50f2/C locus. Absence of this product, with the concomitant presence of the *MSY1* product (1.7–2.7 kb) as an internal PCR control, is diagnostic of deletion of 50f2/C. PCR conditions are: 26 cycles of 95°C 1 min, 66°C 4 min, with primers at 1 µM, in a Perkin Elmer Cetus 4800 thermal cycler.

Established and novel STSs. 50f2/C lies within deletion interval 6E of the Y chromosome (12). Twelve STSs (Research Genetics, Inc.) from intervals 6C–E [sY112, sY144–149 inclusive, and sY152–156 inclusive (16)] and two from the *DAZ* gene [interval 6D; sY254 and sY255 (15)], were tested for co-deletion, using conditions as described, as were three novel STSs designed from the sequence of sub-fragments from two 50f2/C-containing cosmids, M59C4 and M129D3, from the Y-specific cosmid library LL0YNC03 (details of primer sequences and PCR conditions available on request). The latter three PCR assays were uninformative because they amplify identically sized products from other homologous regions of the Y, and these would therefore obscure any absence of products from the deleted region.

Y-specific microsatellites. Microsatellites were amplified in a Perkin Elmer Cetus 9600 thermal cycler using one unlabelled and one ^{33}P-labelled (Amersham) primer, each at 100 nM, and products were separated on 5% denaturing polyacrylamide gels (Sequagel, National Diagnostics) and visualised by autoradiography. The new trinucleotide microsatellites *DYS426* and *DYF371* were isolated by hybridisation selection (N.Fretwell, J.A.L.Armour and M.A.Jobling, manuscript in preparation); primers for *DYF371* amplify four (and rarely, five) Y-chromosomal loci, but a third primer can be used to amplify one of these specifically (*DYS425*). Primer sequences for *DYS426* are: 4D11A 5′–GGTGACAAGACGAGACTTTG-TG-3′; 4D11R 5′-CTCAAAGTATGAAAGCATGACC-3′, and for *DYF371* and *DYS425*: 2E10AL 5′-GTGGGCTGAGA-AATTTCTGG-3′; 2E10R 5′-AGTAATTCTGGAGGTAAA-ATGG-3′; 2E10D1 5′-TGGAGAGAAGAAGAGAGAAAT-3′. PCR conditions, for *DYS390* (1), *DYS426* and *DYF371* (primers 2E10AL and R): 24 cycles of 95°C 20 s, 60°C 45 s, 70°C 50 s; for *DYS425* (primers 2E10D1 and R): as before, but with 58°C annealing temperature; *DYS19*: as described previously (35).

ACKNOWLEDGEMENTS

We thank Ma Kun for the probe MK5, Pieter de Jong for the Y-specific cosmid library LL0YNC03, Nathan Ellis and Michael Hammer for DNA samples from the Y Chromosome Consortium Repository, Adolfo López de Munain for Basque DNAs, Doudja Nafa for Algerian DNAs, Bryan Sykes for Cook Islander DNAs, and Ryk Ward for helpful comments and suggestions. We also thank John Armour, Nourdine Bouzekri and Paul Taylor for assistance and comments on the manuscript. P.J.S. thanks Albert de la Chapelle for support. M.A.J. was funded by an MRC Training Fellowship and a Wellcome Trust Career Development Fellowship, N.F. by an MRC-HGMP studentship, A.P. by the BBSRC, B.B. by the Erasmus programme, and C.T.-S. by the CRC. P.J.S. was funded by the Academy of Finland, the Sigrid Jusélius Foundation, the Finnish Cultural Foundation, the Emil Aaltonen Foundation and the Finnish Medical Society Duodecim.

REFERENCES

1. Jobling,M.A. and Tyler-Smith,C. (1995) Fathers and sons—the Y chromosome and human evolution. *Trends Genet.*, **11**, 449–456.
2. Seielstad,M.T., Hebert,J.M., Lin,A.A., Underhill,P.A., Ibrahim,M., Vollrath,D. and Cavalli-Sforza,L.L. (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.*, **3**, 2159–2161.
3. Hammer,M.F. (1995) A recent common ancestry for human Y chromosomes. *Nature*, **378**, 376–378.
4. Whitfield,L.S., Sulston,J.E. and Goodfellow,P.N. (1995) Sequence variation of the human Y chromosome. *Nature*, **378**, 379–380.
5. Underhill,P.A., Jin,L., Zemans,R., Oefner,P.J. and Cavalli-Sforza,L.L. (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl Acad. Sci. USA*, **93**, 196–200.
6. Hammer,M.F. (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. *Mol. Biol. Evol.*, **11**, 749–761.
7. Casanova,M., Leroy,P., Boucekkine,C., Weissenbach,J., Bishop,C., Fellous,M., Purrello,M., Fiori,G. and Siniscalco,M. (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science*, **230**, 1403–1406.
8. Disteche,C.M., Casanova,M., Saal,H., Friedman,C., Sybert,V., Graham,J., Thuline,H., Page,D.C. and Fellous,M. (1986) Small deletions of the short arm of the Y chromosome in 46,XY females. *Proc. Natl Acad. Sci. USA*, **83**, 7841–7844.
9. Nagafuchi,S., Namiki,M., Nakahori,Y., Kondoh,N., Okuyama,A. and Nakagome,Y. (1993) A minute deletion of the Y chromosome in men with azoospermia. *J. Urol.*, **150**, 1155–1157.
10. Tiepolo,L. and Zuffardi,O. (1976) Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum. Genet.*, **34**, 119–124.

11. Ma,K., Sharkey,A., Kirsch,S., Vogt,P., Keil,R., Hargreave,T.B., McBeath,S. and Chandley,A.C. (1992) Towards the molecular localisation of the AZF locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. *Hum. Mol. Genet.*, **1**, 29–33.
12. Foote,S., Vollrath,D., Hilton,A. and Page,D.C. (1992) The human Y chromosome—overlapping DNA clones spanning the euchromatic region. *Science*, **258**, 60–66.
13. Tyler-Smith,C., Oakey,R.J., Larin,Z., Fisher,R.B., Crocker,M., Affara,N.A., Ferguson-Smith,M.A., Muenke,M., Zuffardi,O. and Jobling,M.A. (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nature Genet.*, **5**, 368–375.
14. Ma,K., Inglis,J.D., Sharkey,A., Bickmore,W.A., Hill,R.E., Prosser,E.J., Speed,R.M., Thomson,E.J., Jobling,M., Taylor,K., Wolfe,J., Cooke,H.J., Hargreave,T.B. and Chandley,A.C. (1993) A Y chromosome gene family with RNA-binding protein homology—candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell*, **75**, 1287–1295.
15. Reijo,R., Lee,T.Y., Salo,P., Alagappan,R., Brown,L.G., Rosenberg,M., Rozen,S., Jaffe,T., Straus,D., Hovatta,O., de la Chapelle,A., Silber,S. and Page,D.C. (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genet.*, **10**, 383–393.
16. 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.
17. Weber,J.L. and Wong,C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **3**, 1123–1128.
18. Vogt,P.H., Edelmann,A., Kirsch,S., Henegariu,O., Hirschmann,P., Kiesewetter,F., Köhn,F.M., Schill,W.B., Farah,S., Ramos,C., Hartmann,M., Hartschuh,W., Meschede,D., Behre,H.M., Castel,A., Nieschlag,E., Weidner,W., Gröne,H.-J., Jung,A., Engel,W. and Haidl,G. (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet.*, **5**, 933–943.
19. Santos,F.R., Gerelsaikhan,T., Munkhtuja,B., Oyunsuren,T., Epplen,J.T. and Pena,S.D.J. (1996) Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. *Hum. Genet.*, **97**, 309–313.
20. Spurdle,A.B. and Jenkins,T. (1994) The specific pDP31 rearrangement polymorphism in Southern African populations. *Hum. Hered.*, **44**, 261–265.
21. Chandley,A.C. and Cooke,H.J. (1994) Human male fertility—Y-linked genes and spermatogenesis. *Hum. Mol. Genet.*, **3**, 1449–1452.
22. Sajantila,A. and Pääbo,S. (1995) Language replacement in Scandinavia. *Nature Genet.*, **11**, 359–360.
23. Lahermo,P., Sajantila,A., Sistonen,P., Lukka,M., Aula,P., Peltonen,L. and Savontaus,M.-L. (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am. J. Hum. Genet.*, **58**, 1309–1322.
24. Uibopuu,V. (1988) *Finnougrierna och deras språk.* Studentlitteratur, Lund, Sweden.
25. Sajantila,A., Salem,A.-H., Savolainen,P., Bauer,K., Gehrig,C. and Pääbo,S. (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc. Natl Acad. Sci. USA*, in press.
26. de la Chapelle,A. (1993) Disease gene mapping in isolated human populations: the example of Finland. *J. Med. Genet.*, **30**, 857–865.
27. Roewer,L., Kayser,M., Dieltjes,P., Nagy,M., Bakker,E., Krawczak,M. and de Knijff,P. (1996) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum. Mol. Genet.*, **5**, 1029–1033.
28. Cooper,K.F., Fisher,R.B. and Tyler-Smith,C. (1992) Structure of the pericentric long arm region of the human Y-chromosome. *J. Mol. Biol.*, **228**, 421–432.
29. Chen,Z.Y., Hendriks,R.W., Jobling,M.A., Powell,J.F., Breakefield,X.O., Sims,K.B. and Craig,I.W. (1992) Isolation and characterization of a candidate gene for Norrie Disease. *Nature Genet.*, **1**, 204–208.
30. Jobling,M.A. (1994) A survey of long-range DNA polymorphisms on the human Y chromosome. *Hum. Mol. Genet.*, **3**, 107–114.
31. Mathias,N., Bayés,M. and Tyler-Smith,C. (1994) Highly informative compound haplotypes for the human Y chromosome. *Hum. Mol. Genet.*, **3**, 115–123.
32. Hammer,M.F. and Ellis,N.A. (1995) Appendix 1: The Y Chromosome Consortium Repository. *YCC Newsletter*, **2**, 8–9.
33. 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.
34. Jobling,M.A., Fretwell,N., Dover,G.A. and Jeffreys,A.J. (1994) Digital coding of human Y chromosomes—MVR-PCR at Y-specific minisatellites. *Cytogenet. Cell Genet.*, **67**, 390.
35. Roewer,L., Arnemann,J., Spurr,N.K., Grzeschik,K.H. and Epplen,J.T. (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.*, **89**, 389–394.