Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, MSY1

Mark A. Jobling*, Evelyne Heyer1, Patrick Dieltjes2 and Peter de Knijff2

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK, 1Laboratoire d'Anthropologie Biologique CNRS UMR152, Musée de L'Homme, 17 place du Trocadero, 75116 Paris, France and 2MGC—Department of Human Genetics, Leiden University, PO Box 9503, 2300 RA Leiden, The Netherlands

Received June 28, 1999; Accepted August 4, 1999

Polymorphic Y-chromosome-specific microsatellites are becoming increasingly used in evolutionary and forensic studies and, in particular, in dating the origins of Y-chromosomal lineages. Previously, haplotyping of Y chromosomes from males belonging to a set of deep-rooting pedigrees was used to estimate a conservative average Y-chromosomal microsatellite mutation rate of 2.1 × 10⁻³ per locus per generation. A number of males showed multiple differences in haplotypes compared with other males within their pedigrees, and these were excluded from the calculation of this estimate, on the grounds that non-paternity was a more probable explanation than multiple mutation within a lineage. Here we reanalyse the pedigrees using an independent highly polymorphic system, the Y-specific minisatellite, MSY1. This supports the hypothesis of non-paternity where more than one microsatellite difference was observed, provides further support for the previously deduced microsatellite mutation rate and throws light on the mutation dynamics of MSY1 itself, suggesting that single-step changes are not the only mode of mutation.

INTRODUCTION

Molecular analysis of the human Y chromosome is playing an increasingly important role in evolutionary (1,2), forensic (3) and genealogical (4) studies. This progress is due to the development of new polymorphic systems on the chromosome, from slowly mutating biallelic markers such as base substitutions (5), to more rapidly mutating and multiallelic microsatellites (6,7) and a single minisatellite (8). In evolutionary studies, the biallelic markers are used to define lineages, and then the multiallelic markers can be used to assay the internal diversity of a lineage, and to estimate the age of the most recent common ancestor of the chromosomes examined (9–11); this relies upon good estimates of the mutation rates of the multiallelic loci. This information is also important if they are to be used in forensic casework (3) or deficiency paternity testing (3,12).

Previously, DNAs from 42 contemporary male members of 12 deep-rooting pedigrees from the Saguenay region of north-east Quebec, Canada, were haplotyped using nine Y-specific microsatellites—one tri-, one penta- and seven tetranucleotide repeats (13); these males represented 257 independent father–son transmissions of the Y chromosome. Four instances of single-step changes in allele length at a single microsatellite locus were observed, and these were assumed to represent mutations; however, as well as these, one male was observed with seven, one with three, and one with two differences in allele length from other pedigree members. While it seems reasonable to assume that the individual with seven changes results from non-paternity, the others are more problematic. Three different mutation rates were presented (13), including and excluding these individuals showing multiple changes. The lowest of these, an average rate of 2.1 × 10⁻³ per locus per generation [95% confidence interval (95% CI) 0.6–4.9 × 10⁻³] has become widely used in evolutionary studies (10,11,14–17).

The locus MSY1 (DYF155S1) (8) is the only known polymorphic Y-specific member of another class of tandem-repeat loci, minisatellites. It consists of an array of ~50–115 repeats of a 25 bp unit which varies in sequence through base substitution. Positions of different variant repeat units can be mapped along the array by minisatellite variant repeat (MVR)-PCR, to generate 'MSY1 codes'. These are extremely variable, and calculations from diversity analysis suggest a high mutation rate, of between 2 and 11% per generation (8).

Here, we re-analyse the deep-rooting pedigrees using this independent highly polymorphic system; this supports the assumption that non-paternity explains the apparent multiple microsatellite mutants, thus strengthening the validity of the 2.1 × 10⁻³ rate, and also provides some preliminary information regarding mutation dynamics at MSY1.

RESULTS

Three-state MSY1 codes, where repeat types designated 1, 3 and 4 are typed, were determined by MVR-PCR for the 42 males from the deep-rooting pedigrees. Occasional repeats within alleles are neither types 1, 3 or 4, and these are designated ‘nulls’, and represented with a ‘0’. The MSY1 codes are shown in Figure 1, together with the microsatellite haplotypes determined previously (13). In the absence of non-paternity or MSY1 mutation we expect the 12 pedigrees to present a total of 12 different codes; we find 21. These are described below, in light of the previous assumptions (13) about non-paternity from the microsatellite data, and of mutation at MSY1 itself. Different repeat units are arranged

*To whom correspondence should be addressed. Tel: +44 116 252 3377; Fax: +44 116 252 3378; Email: maj4@leicester.ac.uk
in blocks along the MSY1 array, and the order of these blocks defines a simpler property of a code, its modular structure. Throughout this discussion, modular structure is referred to as in the following example: ‘1,3,4’—a 5′ block of type 1 repeats, a central block of type 3 repeats and a 3′ block of type 4 repeats.

Below we make qualitative statements about the relative likelihood of males sharing microsatellite haplotypes or MSY1 codes by descent, and this sharing arising through coincidence. In principle these statements could be made quantitative, but this would require knowledge of haplotype frequencies and non-paternity frequency in the Saguenay population, and independent information about mutation rates, none of which we possess. We have therefore chosen not to make these estimates. However, it should be borne in mind that average haplotype frequencies, at least, are low; in European populations a given seven-locus microsatellite haplotype is shared by an average of 1.5 unrelated males (n = 322) (7), while for MSY1 codes the corresponding figure is only 1.2 (n = 160) (8).

MSY1 codes in males representing either non-paternity or multiple microsatellite mutation

Male E. This lineage lies within the large pedigree descending from male 18219, for which an ancestral microsatellite haplotype of 254522252, and an ancestral MSY1 code of (1)5(3)5(4)15 can be inferred. Male E (Fig. 1) shows seven allele length differences in the nine microsatellites typed with respect to the probable ancestral haplotype. The MSY1 code of this individual is also clearly distinct from those in the rest of the pedigree, with a different modular structure (3,1,3,4 versus 1,3,4), a difference of four repeat units in overall array length from that of the presumed ancestral code and differences in the lengths of blocks of type 1 and internal type 3 repeats, strongly supporting the assumption of non-paternity.

Males O, P and Q. Within this pedigree, descending from male 18642, male O shows three microsatellite allele length differences in the nine microsatellites typed with respect to the probable ancestral haplotype.
differences from the pair of males, P and Q, descending from male 20112. This pair does not have identical MSY1 codes: both are 79 repeats in length, belonging to the ‘1,3,4’ modular structural class, with identical numbers of type 1 repeats. However, the block of type 3 repeats is longer by one repeat in male Q than in his uncle, P, and the block of type 4 repeats is correspondingly shorter. We assume here that this is likely to represent MSY1 mutation, rather than non-paternity with coincidental microsatellite haplotype sharing. The mutation results either from two independent single-step mutation events, or, alternatively, a switch of repeat type at the boundary between type 3 and 4 repeats in a single event [this putative single event we refer to as a ‘boundary switch’ (see below)]. Male O has an allele length five repeats shorter than P/Q, and this results from differences in at least two of the three repeat blocks. Taken with the three microsatellite allele length differences in this individual, this suggests that a non-paternity between O and P/Q is likely.

**Males K and L.** There are two microsatellite allele length differences between these two individuals, descending from male 18251; MSY1 also supports the assumption of non-paternity here, showing different modular structures (1,3,4 versus 1,3,4,0,4), a difference of four repeat units in overall array length and differences in the lengths of all comparable repeat blocks.

**MSY1 codes in males representing single microsatellite mutations**

Three pedigrees, descending from males 18255 (within the large pedigree of male 18219), 18614 and 88583 show single aberrant microsatellite allele lengths: interpreted as mutations, these give rise to the conservative average mutation rate estimate of $2.1 \times 10^{-3}$. Both descendants of male 18614 (males I and J) show identical MSY1 codes, as do both descendants of male 88583 (males M and N), thus supporting the inherent assumption of correct paternity for these males, and hence the assumption that the microsatellite allele length changes do indeed represent mutations. One male (male G) in the dichotomous pedigree descending from male 14730 has the presumed ancestral MSY1 code, (1) (3) (3) (4) (3) (4), while the other, male B, has a single repeat unit decrease in the block of type 3 repeats, which represents a probable MSY1 mutation.

**Other MSY1 codes showing differences within pedigrees**

Apart from those discussed above, there are seven further males (A, D, F, G, H, R and S) who show MSY1 codes different from those in other pedigree members, with no concomitant differences in microsatellite haplotypes. These differences can be explained either by mutation at MSY1, or by non-paternity, with coincidental microsatellite haplotype identity. In most cases, we can say little about the exact nature of possible MSY1 mutations because so many father–son transmissions are unobserved in these pedigrees; differences in contemporary members could reflect single or multiple mutation events.

However, the only observed MSY1 difference which can most conservatively be accounted for by a single-step mutation is that seen in male B. Other observed MSY1 differences are either due to multiple simple events within a lineage, or to a single more complex event; males A, D and G/H show the same phenomenon as that observed in males P and Q, a ‘boundary switch’, where an increase in the length of one repeat block is accompanied by an equivalent decrease in the length of an adjacent block, with no change in overall array length. In addition, male F shows a three-repeat increase in the block of type 3 repeats with respect to the presumed ancestral code in nine generations, and males R and S may represent a more complex event(s) including a change in modular structure. The observed distribution of mutations within the pedigrees suggests that complex events may be a more probable explanation than multiple single events; excluding males E, K/L and O, and assuming that males G and H inherit their shared code from a common ancestor, we see seven such events. These can be converted into 16 single-step equivalents (including, with equal weight, a repeat-type switch in males R/S). In the absence of substantial mutation rate heterogeneity between lineages, it is highly unlikely that all 16 single-step mutations would cluster in the way that we observe ($P < 0.005$; exact test).

**DISCUSSION**

MSY1 is a highly polymorphic locus whose informativeness allows a test of the assumptions of non-paternity (or, indeed, true paternity) which were made previously in the study of Y-specific microsatellite mutation rates. Here we have shown that cases where males within a pedigree show more than one microsatellite difference are very likely to be due to non-paternity, and this study thus provides support for the most conservative microsatellite mutation rate of $2.1 \times 10^{-3}$ (13).

Further support for our conclusions comes from another study where MSY1 and Y-specific microsatellites were typed in parallel in deep-rooting pedigrees, in an investigation of the Thomas Jefferson paternity case (4). Here, 61 father–son transmissions of the Y chromosome were surveyed (excluding one lineage where recent non-paternity was evident): only two instances of microsatellite differences within pedigrees were observed, both of which involved single loci within eleven-locus haplotypes, and both of which were single repeat unit changes from the probable ancestral haplotypes. MSY1 codes within these lineages were consistent with true paternity, and in all the pedigrees analysed, two instances of apparent MSY1 mutation were observed, one explicable by a single-step increase within a repeat block, and one by a ‘boundary switch’ event, like the one observed in male A.

However, it must be remembered that the use of MSY1 and the microsatellites together does not provide a truly independent inclusive test of true paternity; because all of these loci are permanently linked, they pass along identical evolutionary trajectories. Thus, within a lineage, Y chromosomes having similar microsatellite haplotypes will tend to have similar MSY1 codes and, if a particular microsatellite haplotype is common in a population or sub-population, MSY1 may not convincingly reveal a non-paternity. However, MSY1 is more likely to provide a good exclusion paternity test, since a discordant microsatellite haplotype will then tend to be accompanied by a discordant MSY1 code. The underlying assumption here is that any genome-wide (or chromosome-wide) elevation in microsatellite mutation rate is unlikely to affect MSY1 in the same way; for example, it is assumed that the discordant microsatellite haplotype and MSY1 code of male E both reflect non-paternity, rather than some simultaneous disturbance in mutation in both systems with a common cause. We believe that such an assumption is justified, given the major differences in repeat unit length, array length,
internal structural diversity, and potential for secondary structure, between MSY1 and the microsatellites. As additional support for this assumption, in diseases such as HNPCC, where microsatellite instability is a characteristic (18), no accompanying disturbance of minisatellite mutation has been observed (19).

We can estimate a mutation rate for MSY1 from these data if we assume that each contemporary pedigree member showing an MSY1 code difference reflects a single, usually complex, mutation event (see Results). We see seven mutations in 224 transmissions, giving a general rate for mutation events, regardless of type, of ~3% (95% CI 1.4–5.8%). Converting these different apparent mutations into 16 single-step equivalents (see Results) gives an effective single-step rate of ~7% (95% CI 2.8–10.4%).

This work strongly suggests that there is complexity in the MSY1 mutation process, and that a single-step model will not be adequate to explain mutation at this locus. This complexity will be addressed more rigorously by studies in sperm DNA, and in pedigree studies where all father–son transmissions can be observed, and where paternities can be verified with non-Y-chromosomal markers (P.G. Taylor, P.C. Patals, T.A. Meitinger and M.A. Jobling, manuscript in preparation).

MATERIALS AND METHODS

DNA samples from 42 males were as described (13). Three-state MVR-PCR (mapping repeat types 1, 3 and 4) at the minisatellite, MSY1, was carried out as described (8).

For MSY1 mutation rates 95% CI limits were estimated by simulation from the binomial distribution (13).

ACKNOWLEDGEMENTS

We thank Matthew Hurles and Chris Tyler-Smith for comments on the manuscript, and Charles Brenner for valuable assistance. M.A.J. is a Wellcome Senior Research Fellow in Basic Biomedical Science, and this work was funded through a Wellcome Research Career Development Fellowship (grant no. 044910).

REFERENCES