Am. J. Hum. Genet. 66:1580–1588, 2000

Characteristics and Frequency of Germline Mutations at Microsatellite Loci from the Human Y Chromosome, as Revealed by Direct Observation in Father/Son Pairs

Manfred Kayser,[1] Lutz Roewer,[2] Minttu Hedman,[3] Lotte Henke,[4] Jürgen Henke,[4] Silke Brauer,[1] Carmen Krüger,[2] Michael Krawczak,[5] Marion Nagy,[2] Tadeusz Dobosz ,[6] Reinhard Szibor,[7] Peter de Knijff,[8] Mark Stoneking,[1] and Antti Sajantila [3]

[1]Max-Planck-Institut für evolutionäre Anthropologie, Leipzig; [2]Institut für Rechtsmedizin, Charité, Humboldt-Universität zu Berlin, Berlin; [3]Department of Forensic Medicine, University of Helsinki, Helsinki; [4]Institut für Blutgruppenforschung, Köln, Germany; [5]University of Wales, College of Medicine, Cardiff; [6]Institute of Forensic Medicine, Medical University Wroclaw, Wroclaw, Poland; [7] Institut für Rechtsmedizin, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany; and [8]Leiden University Medical Center, Leiden, The Netherlands

A number of applications of analysis of human Y-chromosome microsatellite loci to human evolution and forensic science require reliable estimates of the mutation rate and knowledge of the mutational mechanism. We therefore screened a total of 4,999 meioses from father/son pairs with confirmed paternity (probability ⩾99.9%) at 15 Y-chromosomal microsatellite loci and identified 14 mutations. The locus-specific mutation-rate estimates were 0–8.58 × 10^{-3}, and the average mutation rate estimates were 3.17 × 10^{-3} (95% confidence interval [CI] 1.89–4.94 × 10^{-3}) across 8 tetranucleotide microsatellites and 2.80 × 10^{-3} (95% CI 1.72–4.27 × 10^{-3}) across all 15 Y-chromosomal microsatellites studied. Our data show a mutational bias toward length increase, on the basis of observation of more repeat gains than losses (10:4). The data are in almost complete agreement with the stepwise-mutation model, with 13 single-repeat changes and 1 double-repeat change. Sequence analysis revealed that all mutations occurred in uninterrupted homogenous arrays of ⩾11 repeats. We conclude that mutation rates and characteristics of human Y-chromosomal microsatellites are consistent with those of autosomal microsatellites. This indicates that the general mutational mechanism of microsatellites is independent of recombination.

Introduction

Reliable estimates of the mutation rates of Y-chromosomal microsatellites are a prerequisite for dating the origin of Y-chromosomal lineages defined by single-nucleotide polymorphisms (SNPs), as attempted in molecular-anthropology studies (e.g., Zerjal et al. 1997; Bianchi et al. 1998; Lahermo et al. 1999). The knowledge of mutation rates is also necessary for the accurate interpretation of Y-chromosomal microsatellite data in paternity testing and forensic casework (Jobling et al. 1997; Kayser et al. 1997). For autosomal microsatellites, relative mutation rates have been estimated by the application of statistical models to population-genetic data (Di Rienzo et al. 1994; Chakraborty et al. 1997). However, those estimates rely on a number of assumptions, rather than on true data (such as effective population size or population separation time). Thus, absolute mutation-rate estimates as revealed by either direct observations of mutations in families (Weber and Wong 1993; Talbot et al. 1995; Brinkmann et al. 1998; Sajantila et al. 1999) or single sperm cells (Zhang et al. 1994) provide values that are more accurate. In these studies, absolute-mutation-rate estimates in the range of 10^{-2} to 10^{-4} have been reported for autosomal microsatellites.

In contrast to autosomal microsatellites, there are few studies estimating Y-chromosomal microsatellite mutation rates. Heyer et al. (1997) analyzed nine loci in descendants of multigenerational pedigrees, covering 213 meioses, and found four mutations, for an average mutation-rate estimate of 2.1 × 10^{-3}. When multigenerational pedigrees are used, the question of nonpaternity versus mutation cannot be completely resolved, since only remote descendants in the male lineages are analyzed. Thus, the direct analysis of father/son pairs with confirmed paternity is the method of choice for estimation of Y-chromosomal microsatellite mutation rates. In a previous study, we calculated, for the single microsatellite locus DYS19, a rate of 3.19 × 10^{-3} (95% CI 0.41–6.7 × 10^{-3}), on the basis of 626 meioses (Kayser et al. 1997). In the present study, we provide data on 14 germline mutational events from 4,999 meioses at 15 Y-chromosomal microsatellite loci, revealed by

Received December 17, 1999; accepted for publication February 23, 2000; electronically published April 6, 2000.

Address for correspondence and reprints: Dr. Manfred Kayser, Max-Planck-Institut für evolutionäre Anthropologie, Inselstrasse 22, D-04103 Leipzig, Germany. E-mail: kayser@eva.mpg.de

0002-9297/2000/6605-0013$02.00

direct observation, in father/son pairs with confirmed paternity. We have estimated locus-specific and average mutation rates, have determined the sequence of mutated and mutant alleles, and have compared this information to that from previous studies on autosomal microsatellites.

Material and Methods

Samples and Analysis Used for Paternity Testing

Blood samples submitted for routine paternity testing in Germany and Poland were included in this study. Paternity testing was performed by analysis of several types and numbers of genetic markers, including blood-group and enzyme polymorphisms, microsatellites, minisatellites, and/or HLA polymorphisms. Only those father/son pairs for which the biological paternity of the alleged father was confirmed with a probability of ⩾99.9% were used for Y-chromosomal mutation studies. In two father/son pairs, a total of 40 autosomal tetranucleotide microsatellite loci were analyzed. (Detailed information about genetic markers and typing methods used for paternity testing are available on request from the corresponding author.)

Y-Chromosomal DNA Analysis

The Y-chromosomal microsatellites used were DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, YCAIa/b, YCAIIa/b, and DYS413a/b (or YCAIIIa/b), which were typed as described elsewhere (Kayser et al. 1997). Standardization of microsatellite-allele sizing, between different labs, was assured by use of either sequenced allelic ladders or control DNA samples. Mutations were identified as allelic differences between father and son and were confirmed by additional typing and by DNA sequence analysis. Sequence analysis was performed on both DNA strands by use of the SequiTherm™ Cycle Sequencing Kit (Epicentre Technologies) and an A.L.F.express™ automated DNA sequencer (Pharmacia Biotech), after strand separation by the Dynabead system (Dynal). Alternatively, the PRISM™ Ready Reaction Dye Terminator or ABI PRISM™ BigDye Reaction Terminator Cycle Sequencing Kit and an ABI PRISM™ DNA-Sequencer 373 or 377 (PE Biosystems) were used. Published primers (Kayser et al. 1997) were used for PCR and sequencing, except for DYS389 and DYS413. For DYS389, the primers used for PCR and for sequence analysis were as follows: primer 1, 5′-TCATCTGTATTATCTATGTGTG-3′; and primer 2, 5′-CCAGACATTGCCAAGTGTTACTTG-3′. The PCR conditions as follows: 25-μl volume with 2.5 μl of 10 × PCR buffer (Promega), 0.5 μl of 2.5 mM of each dNTP (Boehringer-Mannheim), 1.5 μl of 25 mM $MgCl_2$, 1 μl of 10 μM of each primer, 1 U of *Taq* polymerase (Promega), and 50–200 ng of DNA; and 30 cycles of 30 s at 94°C, 30 s at 54°C, and 45 s at 72°C, in a Thermocycler PTC 100 (MJ Research). For PCR and sequencing of locus DYS413, the following primers, based on the published sequence (Malaspina et al. 1997), were used to amplify the Y chromosome–specific but not the X chromosome–specific loci: primer 1, 5′-AATGTGTGAGCCAATTGTTTAGAA-3′; and primer 2, 5′-TCAGAGAAGGAGAAACTAAACCAAA-3′. The PCR conditions were the same as those used for DYS389, except that the annealing temperature was 61°C. DNA sequence analysis for DYS413 was done either directly from PCR products or after cloning of the PCR product by the TOPO TA Cloning Kit (Invitrogene).

Statistical Analysis

The probability of paternity was calculated separately for single-locus systems (Essen-Möller 1938) and for multilocus DNA-fingerprint data (Krawczak and Bockel 1992). Mutation rates were estimated as the number of mutations, divided by the number of meioses analyzed. The 95% Poisson confidence interval (CI) was calculated as described elsewhere (Heyer et al. 1997). Binomial probabilities comparing the number of repeat gains versus loss, the number of observed mutation events in tetranucleotide- versus dinucleotide-repeat microsatellites, and different mutation-frequency estimates were calculated by use of an exact test comparing two binomial probabilities, as implemented in StatXact (Cytel Software). In each case, the binomial probability indicates whether the different values deviate significantly from the null hypothesis of a 1:1 ratio. Statistical significance of correlation coefficients from comparison of the number of repeats and mutation rates, as well as length differences in mutated and nonmutated alleles, were calculated by a *t*-test. Statistical significance for the number of nonmutated and mutated alleles being either (*a*) longer than or equal to 11 repeats or (*b*) shorter than 11 repeats was calculated by a χ^2 test. The probability ($P_{k,n}$) that, during a single meiosis, $\geqslant k$ of n loci undergo a mutation with mutation rate μ is that of a binomial distribution,

$$P_{k,n} = \sum_{j=k}^{n} \binom{n}{j} \mu^j (1-\mu)^{n-j} = 1 - \sum_{j=0}^{k-1} \binom{n}{j} \mu^j (1-\mu)^{n-j} ,$$

whereas the probability ($P_{m,k,n}$) that none of m meioses shows a mutation at $\geqslant k$ of n loci is $P_{m,k,n} = (1 - P_{k,n})^m$. The ages of the fathers with mutations and of those without mutations were compared by a nonparametric Mann-Whitney U rank-sum test.

Results

Mutation-Rate Estimates for Y-Chromosomal Microsatellites

We analyzed a total of 4,999 male germline transmission events, in confirmed father/son pairs, at eight Y-chromosomal tetranucleotide microsatellites (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS393, DYS385a/b), one trinucleotide microsatellite (DYS392), and six dinucleotide microsatellites (YCAIa/b, YCAIIa/b, and DYS413a/b). In this survey we observed 14 mutations, including 5 that we had reported elsewhere (Kayser et al. 1997) (table 1). In the families in which mutations were identified, paternity was assured, with probabilities of 99.999998%–99.9999999999%, by analysis of 10–33 genetic markers (table 1). The locus-specific mutation-rate estimates were 0–8.58×10^{-3} (table 2). The average tetranucleotide mutation rate across 8 loci was 3.17×10^{-3} (95% CI 1.89–4.94×10^{-3}); the average dinucleotide mutation rate across 6 loci was 2.04×10^{-3} (95% CI 0.06–10.93×10^{-3}); and the overall average mutation rate across all 15 loci studied was 2.80×10^{-3} (95% CI 1.72–4.27×10^{-3}) (table 2).

Characteristics of Y-Chromosomal Microsatellite Mutations

Of the 14 mutations identified, a 10:4 ratio of gains versus losses of microsatellite repeats ($P = .09$), as well as a 13:1 ratio of mutations involving a single-repeat change versus a double-repeat change, was observed (table 1).

For further characterization of the mutations, the father/son nucleotide sequences involved in each mutation event were determined. In all 14 cases, the premutation allele size consisted of uninterrupted arrays of ⩾11 repeats that are completely homogeneous in size and sequence (table 1). However, when, across all tetranucleotide loci, the number of nonmutated and mutated alleles with ⩾11 homogeneous repeats is compared with those with <11 homogeneous repeats, the difference is not statistically significant. Of the loci analyzed, at least seven (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, and DYS391) are compound microsatellites, containing more than one type of repeat. The majority of mutations (13 of 14) were observed at these compound microsatellites (tables 1 and 2) and occurred al-

Table 1

Interpretation of 14 Mutations Observed at Y-Chromosomal Microsatellite Loci

Locus	Father	Son	Repetitive Sequence Structure→Mutation[a]	Type	No. of Steps	Paternity Probabilities SLS (No. of Genetic Markers Used) / MLS[b]
DYS19[c]	14	15	…(CTAT)$_{11\rightarrow12}$(CTAC)$_1$(CTAT)$_3$…	Gain	One	99.999999998 (27) / 99.99999
DYS19[c]	14	14	…(CTAT)$_{11\rightarrow12}$(CTAC)$_1$(CTAT)$_3$…	Gain	One	99.999999 (16)
DYS389I	11	10	…(CTGT)$_4$(CTAT)$_{12}$…(CTGT)$_3$(CTAT)$_{11\rightarrow10}$…	Loss	One	99.9999996 (15)
DYS389II[c]	16[d]	17[d]	…(CTGT)$_5$(CTAT)$_{11\rightarrow12}$…(CTGT)$_3$(CTAT)$_9$…*	Gain	One	99.999999999 (12) / 99.999
DYS389II	17[d]	18[d]	…(CTGT)$_6$(CTAT)$_{11\rightarrow12}$…(CTGT)$_3$(CTAT)$_{10}$…	Gain	One	99.9999999999 (33)
DYS390[c]	26	25	…(CTAT)$_2$(CTGT)$_8$(CTAT)$_{13\rightarrow12}$(CTGT)$_1$(CTAT)$_4$…*	Loss	One	99.999999999 (12) / 99.999
DYS390	24	25	…(CTAT)$_2$(CTGT)$_8$(CTAT)$_{11\rightarrow12}$(CTGT)$_1$(CTAT)$_4$…†	Gain	One	99.9999993 (25) / 99.9998
DYS390	26	25	…(CTAT)$_2$(CTGT)$_8$(CTAT)$_{13\rightarrow12}$(CTGT)$_1$(CTAT)$_4$…	Loss	One	99.999999 (15)
DYS390	26	27	…(CTAT)$_2$(CTGT)$_8$(CTAT)$_{13\rightarrow14}$(CTGT)$_1$(CTAT)$_4$…	Gain	One	99.99999993 (15)
DYS391	11	12	…(CTGT)$_3$(CTAT)$_{11\rightarrow12}$…	Gain	One	99.999999 (25)
DYS391	11	12	…(CTGT)$_3$(CTAT)$_{11\rightarrow12}$…	Gain	One	99.9999995 (13)
DYS413a/b	21–22	21–20	…(CA)$_{21}$ …(CA)$_{22\rightarrow20}$…†,e	Loss	Two	99.9999993 (25) / 99.9998
DYS385a/b[c]	12–16	12–17	…(AAGG)$_6$(GAAA)$_{12}$…(AAGG)$_6$(GAAA)$_{16\rightarrow17}$e,f…	Gain	One	99.999998 (10)
DYS385a/b	11–14	11–15	…(AAGG)$_6$(GAAA)$_{11}$ …(AAGG)$_6$(GAAA)$_{14\rightarrow15}$e,f…	Gain	One	99.99999993 (15)

[a] Repetitive sequence structure gives the total number of repeats; mutated repeats are denoted by arrows (→), and an ellipsis (…) denotes the interspersed and surrounding single-copy sequence; the asterisks (*) denote two mutations found in one father/son pair, and the daggers (†) denote two mutations found in another father/son pair. Allele nomenclature according to Kayser et al. (1997) is indicated by the underlined repeats.

[b] SLS = single-locus systems such as blood group/enzyme polymorphisms, HLA polymorphisms, microsatellites, and VNTRs; MLS = multilocus systems/multilocus DNA fingerprint analysis.

[c] Mutations that we have reported elsewhere (Kayser et al. 1997) but whose mutational characteristics, including sequences, have been determined in the present study.

[d] Excludes DYS389I.

[e] The mutational explanation given here and in the text assumes that alleles found in the father and the son are identical by descent.

[f] Not confirmed by sequence analysis, because of lack of DNA; the sequence was deduced from the fragment length, as described in the text.

ways in the longest array of homogeneous repeats (table 1).

Comparing, for all seven mutated microsatellites, the mean number of uninterrupted repeats of all mutant alleles (postmutation allele size) versus that of all nonmutated alleles, we found that the mutant alleles are longer at four loci (DYS385a/b, DYS389II, DYS390, and DYS391), the same mean length at two loci (DYS19 and DYS389I), and shorter at one locus (DYS413a/b) (fig. 1).

The mutations always occurred in either the most common alleles or alleles longer than the most common one (fig. 2). A comparison of the mean repeat length of nonmutated and mutated alleles (premutation allele size), across the tetranucleotide microsatellite loci analyzed here, revealed that the mutated alleles were always longer, when both homogeneous and total repeats (with and without DYS390) were considered, but the differences were not statistically significant. Comparison of the mean number of repeats (premutation allele size) versus the mutation rate for each locus revealed no significant correlation when only the number of homogeneous repeats was considered but revealed a positive significant correlation when the total number of tetranucleotide repeats was considered (correlation coefficient .73). However, when DYS390 was removed from the analysis, the correlation was no longer statistically significant.

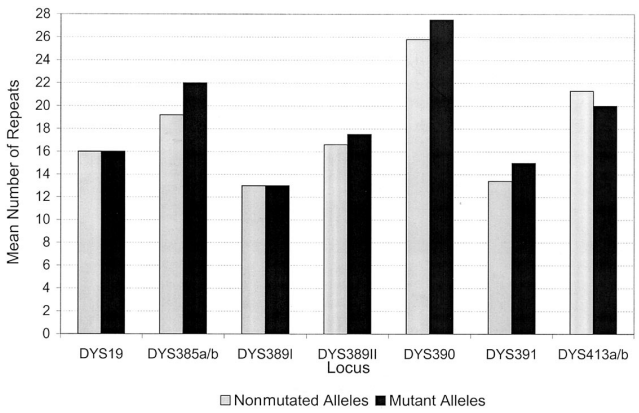

Figure 1 Comparison of mean total number of repeats, between nonmutated alleles and mutant alleles (postmutation allele size), for Y-chromosomal microsatellite loci.

For the 12 males with a mutation, the father's age at the birth of the son was 17–40 years, with an average age of 25.9 ± 7.0 years; for the fathers without a mu-

Table 2

Y-Chromosomal Microsatellite Mutation Rates, as Revealed by Father/Son–Pair Analysis

Marker	Gene Diversity[a]	Allele Size[a] (bp)	Variable Repeat(s)	No. of Repeats[b]		No. of Mutations[c]	No. of Meioses[c]	Mutation Rate [95% CI[d]] ($\mu \times 10^{-3}$)
				Homogeneous[e]	Total[f]			
DYS19	.74	178–206	CTAT/CTAC	12.0	16.0	2	996	2.01 [.26–6.82]
DYS385a/b	.88	356–408	AAGG/GAAA[g]	13.2	19.2	2	952	2.1 [.27–7.15]
DYS389I	.81	247–255	CTGT/CTAT	10.0	13.0	1	425	2.35 [.06–12.5]
DYS389II	.63[h]	357–381	CTGT/CTAT[h]	…[i]	16.6[h]	2[h]	425[h]	4.71 [.6–16.0][h]
DYS390	.74	199–223	CTGT/CTAT	10.8	25.8	4	466	8.58 [2.54–20.3]
DYS391	.49	279–291	CTGT/CTAT	10.4	13.4	2	415	4.82 [.61–16.4]
DYS392	.55	233–260	ATT	11.7	11.7	0	415	0 [7.15]
DYS393	.39	120–136	GATA	13.1	13.1	0	415	0 [7.15]
YCAIa/b	.10	124–132	CA	NS	NS	0	150	
YCAIIa/b	.67	147–165	CA	19.6	19.6	0	240	2.04[j] [.06–10.93]
DYS413a/b	.87	192–204	CA	21.3	21.3	1	100	
Overall						14	4,999	2.80 [1.72–4.27]

[a] Values are from the present study. Combined values are given for the duplicated loci DYS385a/b, YCAIa/b, YCAIIa/b, and DYS413a/b, since unequivocal allele-locus assignment is not possible.

[b] NS = not sequenced.

[c] Includes data that we have reported elsewhere (Kayser et al. 1997).

[d] 95% Poisson CI of mutation rate. (For instances in which a mutation has been observed, both the 97.5% upper CI and the 97.5% lower CI are given; for instances in which no mutation has been observed, only the 95% upper CI is given.)

[e] Mean total number of uninterrupted homogeneous repeats. (If the locus consists of more than one repeat, the length of the longest homogeneous array is given.)

[f] Mean total number of uninterrupted total repeats.

[g] Sequence information is from Schneider et al. (1998).

[h] Excludes DYS389I.

[i] DYS389II consists of two variable homogeneous repeats that are not distinguishable by fragment-length analysis.

[j] Average mutation rate across all six dinucleotide loci is given, because of the low number of meioses studied at single loci.

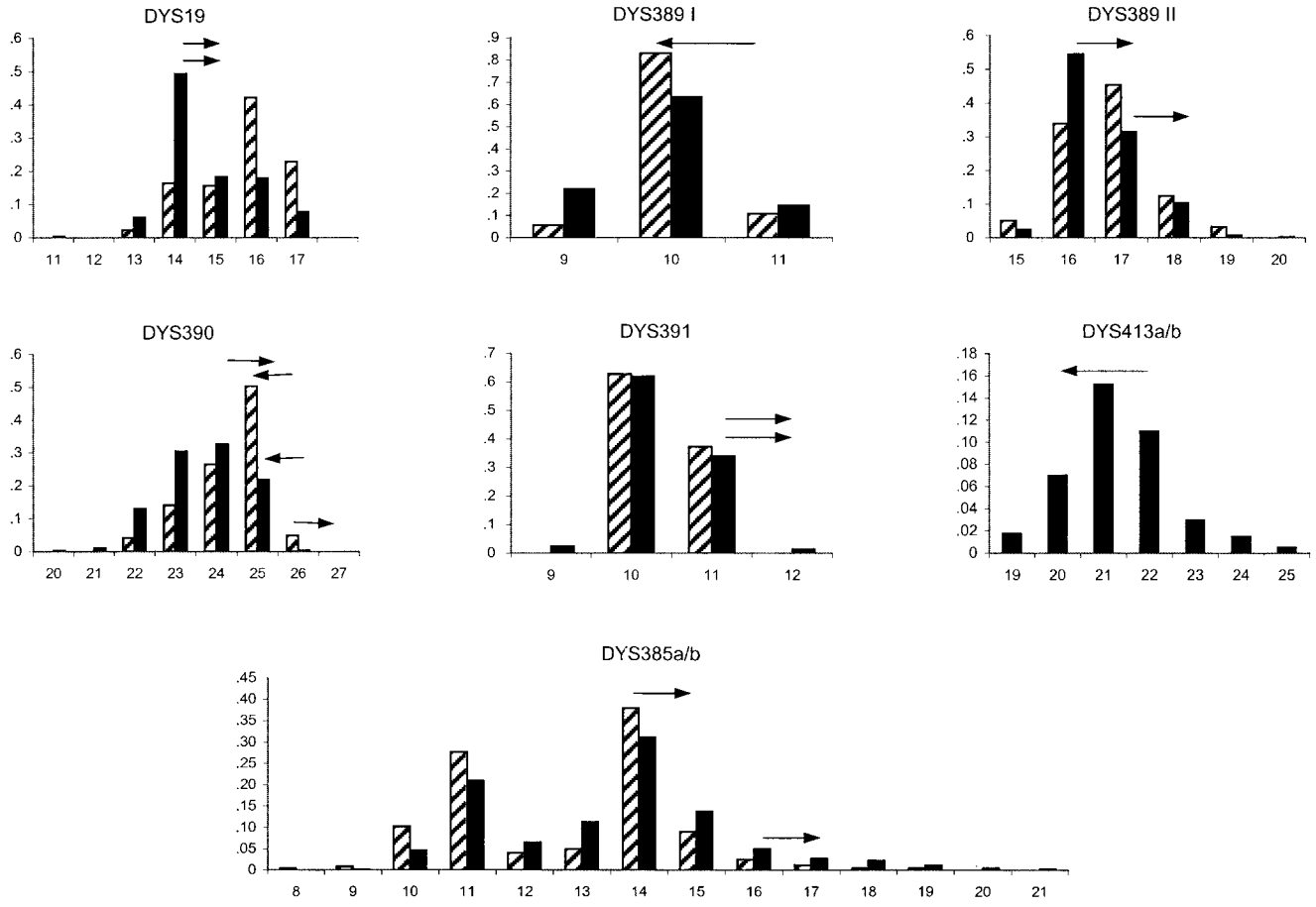

Figure 2 Allele frequencies for Y-chromosomal microsatellites at which mutations were observed. Hatched bars denote a Polish population sample (n = 121), and black bars denote a German population sample (n = 279), which were used for family analysis. Arrows indicate mutational events. The beginning and end of the arrow is centered over the relevant population. For the duplicated loci DYS413a/b and DYS385a/b, combined allele frequencies are given; for DYS413a/b, allele frequencies are based on data for 77 Germans. The *Y*-axis indicates allele frequencies, and the *X*-axis indicates alleles, in terms of number of repeats (nomenclature is according to Kayser et al. [1997]—with the exception of DYS389II, which now excludes variability at DYS389I).

tation, the average age at the birth of the son was 28.9 ± 8.4 years, which was calculated from data on 351 fathers and which covers 66.7% of all the meioses studied. Both groups deviated from a normal distribution; a nonparametric test showed that the age difference between the groups was not significant (P = .215).

Multiple Y-Chromosomal Microsatellite Mutations within the Same Germline Transmissions

A total of 415 father/son pairs were analyzed at the same nine loci (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393). In one father/son pair, mutations at two loci (DYS390 and DYS389) were identified (table 1). The probability P of two mutations in one meiosis, when 415 meioses are screened for these nine loci (which have an average mutation rate of 2.88×10^{-3}), is P = .11. Thus, these two mutations within a single transmission are not unexpected, given our mutation-rate estimate. In another father/son pair, mutations at two loci (DYS390 and DYS413) were found when an additional six dinucleotide microsatellites (YCAIa/b, YCAIIa/b, and DYS413a/b) in a subset of 50 father/son pairs were analyzed (table 1). Here the probability is just barely statistically significant (P = .04); however, the low number of meioses typed for the entire set of 15 markers makes meaningful statistical interpretation critical.

To investigate whether there is a general microsatellite instability in the two families in each of which there were two mutations observed, we analyzed 40 autosomal microsatellites from around the genome. No mutations were observed, whereas the chance of at least

one mutation at those 40 loci would be >95% if the autosomal microsatellite mutation rate were to be elevated by a factor of ⩾25.

Discussion

In this study we used family data covering 4,999 meioses obtained at 15 loci, to estimate Y-chromosomal microsatellite mutation rates. The resulting figures were 0–8.58×10^{-3} for locus-specific rates, 3.17×10^{-3} (95% CI 1.89–4.94×10^{-3}) across the 8 tetranucleotide loci, and 2.80×10^{-3} (95% CI 1.72–4.27×10^{-3}) across all 15 loci studied. For a subset of seven of the loci used here, Heyer et al. (1997) analyzed descendants of 12 multigenerational pedigrees that span ⩽10 generations and that cover 213 meioses. They identified four mutations and estimated the average mutation rates to be 2.0×10^{-3} (95% CI 0.05–5.5×10^{-3}) and 2.1×10^{-3} (95% CI 0.6–4.9×10^{-3}), across nine tri-, tetra-, and pentanucleotide loci studied. Since paternity cannot be established within such multigenerational pedigrees, descendants with more than one allelic difference at Y-chromosomal microsatellites were assumed to reflect nonpaternity and therefore were excluded. Further indication of nonpaternity of those descendants has been reported recently in Jobling et al.'s (1999) analysis of the Y-chromosomal minisatellite MSY1. Our average mutation-rate estimate does not differ significantly from Heyer et al.'s average estimate ($P = .55$), when the same loci are considered.

A second study, by Bianchi et al. (1998), estimated mutation rates for seven of the Y microsatellites used here, via direct observation using cell lines from families. A disadvantage of the use of cell lines for mutation studies is the difficulty in differentiating between somatic mutations produced by the cell-line propagation and true germline mutations that have occurred in the donor individual. Bianchi et al. (1998) observed two mutations but attributed these to somatic events and concluded that no germline mutations occurred in 1,743 meioses. This result is not significantly different from the average mutation rate of 3.09×10^{-3} that we obtained for the same loci ($P = .22$).

When, for the same nine Y-chromosomal microsatellites (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393), all data from the present study and from the Heyer et al. (1997) and Bianchi et al. (1998) studies are pooled, the overall average mutation-rate estimate is 2.08×10^{-3} (95% CI 1.36–3.02×10^{-3}), on the basis of 8,169 meioses, and the average tetranucleotide microsatellite mutation-rate estimate across eight loci is 2.19×10^{-3} (95% CI 1.41–3.23×10^{-3}), on the basis of 7,292 meioses.

Comparison of Mutational Characteristics at Y-Chromosomal and Autosomal Microsatellites

Recently, three studies have reported autosomal microsatellite mutation rate estimates based on large family data sets. The average rates are 2.1×10^{-3}, on the basis of 10,844 meioses at 9 loci (Brinkmann et al. 1998); 2.7×10^{-3}, on the basis of 6,228 meioses at 13 loci (Henke and Henke 1999); and 0.6×10^{-3} on the basis of 16.455 meioses at 5 loci (Sajantila et al. 1999). Comparison of these average mutation rates for autosomal microsatellites to the 2.8×10^{-3} average mutation rate for Y-chromosomal microsatellite, which has been observed in the present study, reveals no significant differences ($P = .15$ for largest difference to the estimate from Sajantila et al.).

For the Y-chromosomal microsatellites studied here, we obtained a higher average mutation rate for tetranucleotide versus dinucleotide microsatellites, but the difference is not significant ($P = .73$). However, more loci (8 vs. 6) and many more meioses (4,094 vs. 490) were studied at tetranucleotide versus dinucleotide microsatellites. Thus, reliable conclusions concerning the size of the repeat and the mutation rate at Y-chromosomal microsatellites are premature. Higher mutation rates for tetra- versus dinucleotide repeats also were observed for autosomal microsatellites, by family analysis (Weber and Wong 1993), but the opposite was revealed by an indirect approach studying microsatellite variability in various populations (Chakraborty et al. 1997). However, the latter approach might be influenced by differences in parameters other than the mutation rate, such as range constraints.

In this study, most of the Y-chromosomal mutations were observed at compound microsatellites (i.e., those containing more than one type of repeat), and all loci with no mutations were simple microsatellites (i.e., contained only one type of repeat). Within the tetranucleotide loci, all 13 mutation were found at the compound microsatellites ($\mu = 3.53 \times 10^{-3}$, 95% CI 2.11–$5.49 \times 10^{-3}$), and no mutation was observed at simple microsatellites (95% upper CI 8.86×10^{-3}); but the difference is not statistically significant ($P = .43$). However, more loci (7 vs. 1) and many more meioses (3,679 vs. 415) were studied at compound than at simple tetranucleotide microsatellites.

Of the 14 Y-chromosomal microsatellite mutations described here, 12 were confirmed by sequence analysis. For the two mutations, at DYS385a/b, that, because of the lack of DNA, could not be confirmed by sequence analysis, the number of repeats was deduced from the fragment lengths, by use of sequenced allelic ladders and the repeat nomenclature from Schneider et al. (1998). All of the mutations occurred at uninterrupted arrays of ⩾11 complete homogeneous repeats. Sequence analysis

of the four Y-chromosomal microsatellite mutations reported by Heyer et al. (1997) revealed that those mutations occurred in uninterrupted arrays of ⩾10 complete homogeneous repeats (data not shown). Similarly, all of the 23 autosomal microsatellite mutations identified by Brinkmann et al. (1998) were observed in uninterrupted arrays of >10 complete homogeneous repeats. On the basis of the combined data, it can be postulated that the probability that human autosomal and Y-chromosomal microsatellites will mutate increases when the length of the array of uninterrupted homogeneous repeats accumulates, with the threshold possibly being ~10 repeats. However, more data are needed to rule out the possibility that this might simply reflect that alleles with <10 homogeneous repeats are rare.

Various tests showed that the number of homogenous repeats was not correlated with mutation rate, either within or across the tetranucleotide loci analyzed here. Since all of the mutations at tetranucleotide microsatellites occurred at compound loci, we also examined the relationship between the total number of repeats and the mutation rate. A significantly positive correlation was found, but elimination of DYS390 (which has both the highest number of total tetranucleotide repeats and the highest mutation rate) also eliminated the significance of the correlation. For the homogenous repeats, the mean length of DYS390 was within the range of all other tetranucleotide loci studied. Thus, the higher mutation rate of DYS390 could be caused by the long stretch of total tetranucleotide repeats, but it also could be due to other factors. For autosomal microsatellites, higher mutation rates were found at compound microsatellites, but these also showed the highest number of homogeneous repeats, and a positive correlation between the mean length of homogenous repeats and the mutation rate was observed (Brinkmann et al. 1998). Increased mutation rates for longer autosomal microsatellite alleles have been reported on the basis of direct observations in yeast (Wierdl et al. 1997), *Drosophila* (Schlötterer et al. 1998), and, indirectly, from phylogenetic analysis in both human (Jin et al. 1996) and *Drosophila* (Goldstein and Clark 1995) populations.

Our data on Y-chromosomal microsatellites provide evidence for a mutational bias in favor of microsatellite expansion: there is a 2.5-fold-higher number of mutations resulting in a repeat gain versus a repeat loss; and, at four of seven loci, a higher mean repeat number in mutant alleles versus nonmutated alleles. Evidence suggestive of length increase also has been detected, by application of a Markov chain Monte Carlo analysis to haplotype data on five of the Y-chromosomal microsatellites used here (Cooper et al. 1999), and directional evolution has been postulated on the basis of comparison of autosomal microsatellites in human and primates (Amos and Rubinsztein 1996; Amos et al. 1996), although ascertainment bias may account for the latter (Ellegren et al. 1995; Cooper et al. 1998). However, two large family studies of autosomal microsatellites do show a slightly higher (but not significant) number of losses versus gains of repeats due to mutation (Brinkmann et al. 1998; Sajantila et al. 1999).

Among the 14 Y-chromosomal microsatellite mutations, we observed 13 single-step changes and 1 double-step change (DYS413a/b). Most mutations at autosomal microsatellites also have been found to be single-step mutations (Weber and Wong 1993; Amos et al. 1996; Brinkmann et al. 1998). These data are consistent with the stepwise-mutation model (Ohta and Kimura 1973) for microsatellite evolution (Valdes 1993), which assumes that mutations cause an increase or decrease, by 1, in repeat number. Our data do not provide direct proof that DNA-strand or polymerase slippage during replication is the major source of microsatellite mutations (Byrd et al. 1965; Schlötterer and Tautz 1992), but the bias in favor of single-repeat changes, as observed here, does support this hypothesis.

Overall, Y-chromosomal microsatellite mutational characteristics and mutation-rate estimates obtained from father/son-pair analysis are in agreement with family data from autosomal microsatellites. This represents further evidence that the general mutational mechanism of microsatellites is likely to be independent of recombination, an inference that also has been suggested elsewhere (Heyer et al. 1997).

Consequences for Forensic and Anthropological Studies

Y-chromosomal microsatellites are especially useful in paternity testing of deficiency cases, which involve a deceased alleged father and a male child, because characterization of the genotype of the deceased alleged father can be replaced by the analysis of any of his living paternal relatives (Kayser et al. 1998). Mutation-rate estimates are necessary in such paternity testing, to assess the possibility of a potential false exclusion. In particular, multiple mutations within a single germline transmission, as has been observed in two instances here, can be misinterpreted as false exclusion of paternity. Although observations of multiple mutations are rare in the literature, they are occasionally observed in paternity-testing laboratories. To the best of our knowledge, only one study has reported multiple mutations in a single case, and that involved 2 of 12 autosomal microsatellites (Gunn et al. 1997). One reason why this phenomenon is rarely recognized in mutation surveys might be that screening for microsatellite mutations is often performed with pooling of single-locus data from a large number of families, rather than by analysis of a large number of loci in the same families. Clearly, only the latter ap-

proach allows the detection of multiple mutations. Thus, a clear consequence of observations of multiple mutations in the same germline is the necessity to define more carefully the criteria for exclusions in paternity testing. The definition should take into account the number and types of loci used, as well as the known mutation-rate estimates for the loci.

Another mutational feature of Y-chromosomal microsatellites that has consequences for data interpretation in forensic-case analysis is the duplication or triplication of a larger Y-chromosomal region that includes the microsatellite locus, followed by a change in the number of repeats within the microsatellite. In the present study, we observed one father/son pair with three instead of two male-specific alleles at DYS385, and additional cases of multiple alleles, at DYS19, DYS390 and DYS391, have been reported elsewhere (Santos et al. 1996*a,* 1996*b;* Kayser et al. 1997; Redd et al. 1997; L.R., unpublished data). In forensic-case analysis, additional alleles are normally interpreted as mixed profiles and, thus, will lead to incorrect conclusions. Therefore, knowledge of the frequency of such events is of interest for forensic laboratories. At DYS19, we have counted nine duplications in 7,772 individuals (Kayser et al. 1997; M. Thomas and N. Bradman, personal communication), revealing a frequency of 0.12% (which does not include those samples in which a common ancestry of the duplication has been revealed by haplotype studies). However, recent duplication events in which the number of repeats is still the same at both loci remain undetected in fragment-length analysis. These observations might reflect the more general tendency of Y-chromosomal DNA to accumulate various types of repetitive sequences—as much as 40% of the euchromatic region (Foote et al. 1992)—which has been explained by the lack of recombination (Charlesworth et al. 1994).

In anthropological studies using Y- chromosome diversity, there is a growing interest in the application of microsatellite variability to the dating of male lineages that are identified by SNPs, in order to infer population history (Zerjal et al. 1997; Bianchi et al. 1998; Lahermo et al. 1999; Ruiz-Linares et al. 1999). These dating methods depend on reliable estimates of locus-specific mutation rates, and the resulting dates can change drastically when unreliable mutation-rate estimates are used, as has been shown by Heyer et al. (1997) using the data of Underhill et al. (1996).

To conclude, this study has presented the largest extant data set of family material used for estimation of mutation rates at Y-chromosomal microsatellites. The major conclusion is that mutation-rate estimates and mutational characteristics of microsatellites are similar in the male sex chromosome versus the autosomes, which suggests that the mutational mechanism for microsatellites in general is independent from recombination. The data also have consequences for forensic casework and molecular anthropology. However, in order to further increase the reliability of the mutation-rate estimates and better understand the underlying mutational mechanism and the relationship between the mutation rate and repeat size, homogeneity, and number, more family analyses or large-scale single-sperm analysis at Y-chromosomal microsatellites should be performed.

Acknowledgments

We thank Chris Tyler-Smith, Matti Lukka, Peter Donnelly, and Svante Pääbo for fruitful discussions and two anonymous reviewers for helpful comments. We are grateful to Volko Ebeling for DNA sequencing of HLA-DRB1 alleles and to Marion Meschkat for archive help. We also thank Neil Bradman and Mark Thomas for sharing data prior to publication. Peter Nürnberg is gratefully acknowledged for providing PCR primers and PCR conditions for autosomal microsatellites.

References

Amos W, Rubinsztein DC (1996) Microsatellites are subject to directional evolution. Nat Genet 12:13–14

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. Nat Genet 13:390–391

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Margnac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, et al (1998) Characterisation of ancestral and derived Y-chromosome haplotypes of new world native populations. Am J Hum Genet 63:1862–1871

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of tandem repeat. Am J Hum Genet 62:1408–1415

Byrd C, Ohtsuka E, Moon MW, Khorana HG (1965) Synthetic deoxyribo-oligonucleotides as templates for DNA polymerase of Escherichia coli: new DNA-like polymers containing repeating nucleotide sequences. Proc Natl Acad Sci USA 53: 79–86

Chakraborty R, Kimmel M, Stivers D, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94: 1041–1046

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eucaryotes. Nature 371:215–220

Cooper G, Burroughs NJ, Rand DA, Rubinsztein DC, Amos W (1999) Markov chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation. Proc Natl Acad Sci USA 96:11916–11921

Cooper G, Rubinsztein DC, Amos W (1998) Ascertainment bias cannot entirely account for human microsatellites being longer than their chimpanzee homologues. Hum Mol Genet 7:1425–1429

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence

repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias? Nat Genet 11:360–362

Essen-Möller E (1938) Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis: theoretische Grundlagen. Mitt Anthropol Ges Wien 68:9–53

Foote S, Vollrath D, Hilton A, Page DC (1992) The human Y chromosome: overlapping DNA clones spanning the euchromatic region. Science 258:60–66

Goldstein DB, Clark AG (1995) Microsatellite variation in North American populations of Drosophila melanogaster. Nucleic Acids Res 23:3882–3886

Gunn PR, Trueman K, Stapleton P, Klarkowski DB (1997) DNA analysis in disputed parentage: the occurrence of two apparently false exclusions of paternity, both at short tandem repeat (STR) loci, in one child. Electrophoresis 18: 1650–1652

Henke L, Henke J (1999) Mutation rate in human microsatellites. Am J Hum Genet 64:1473

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803

Jin L, Macaubas C, Hallmayer J, Kimura A, Mignot E (1996) Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. Proc Natl Acad Sci USA 93: 15285–15288

Jobling MA, Heyer E, Dieltjes P, de Knijff P (1999) Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, MSY1. Hum Mol Genet 8:2117–2120

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124

Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133, 141–149

Kayser M, Krüger C, Nagy M, Geserick G, Roewer L (1998) Y-chromosomal DNA-analysis in paternity testing: experiences and recommendations. In: Olaisen B, Brinkmann B, Lincoln PJ (eds) Progress in forensic genetics 7. Elsevier Science, Amsterdam, pp 494–496

Krawczak M, Bockel A (1992) A genetic factor model for the statistical analysis of multilocus DNA fingerprints. Electrophoresis 13:10–17

Lahermo P, Savontaus ML, Sistonen P, Beres J, de Knijff P, Aula P, Sajantila A (1999) Y-chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. Eur J Hum Genet 7:447–458

Malaspina P, Ciminelli BM, Viggiano L, Jodice C, Cruciani F, Santolamazza P, Sellito D, et al (1997) Characterization of a small family (CAIII) of microsatellite-containing sequences with X-Y homology. J Mol Evol 44:652–659

Ohta T, Kimura M (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. Genet Res 22:201–204

Redd AJ, Clifford SL, Stoneking M (1997) Multiplex DNA typing of short-tandem-repeat loci on the Y chromosome. Biol Chem 378:923–927

Ruiz-Linares A, Ortiz-Barrientos D, Figueroa M, Mesa N, Munera JG, Bedoya G, Velez ID, et al (1999) Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. Proc Natl Acad Sci USA 96: 6312–6317

Sajantila A, Lukka M, Syvänen A-C (1999) Experimentally observed germline mutations at human micro- and minisatellite loci. Eur J Hum Genet 7:263–266

Santos FR, Gerelsaikhan T, Munkhtuja B, Oyunsuren T, Epplen JT, Pena SDJ (1996*a*) Geographic differences in the allelic frequencies of the human Y-linked tetranucleotide polymorphism DYS19. Hum Genet 97:39–313

Santos FR, Rodriguez-Delfin L, Pena SDJ, Moore J, Weiss KM (1996*b*) North and South Amerindians may have the same major founder Y chromosome haplotype. Am J Hum Genet 58:1369–1370

Schlötterer C, Ritter R, Harr B, Brem G (1998) High mutation rate of a long microsatellite allele in Drosophila melanogaster provides evidence for allele-specific mutation rates. Mol Biol Evol 15:1269–1274

Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C (1998) Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. Forensic Sci Int 97: 61–70

Talbot CC, Avramopoulos D, Gerken S, Chakravarti A, Armour JA, Matsunami N, White R, et al (1995) The tetranucleotide repeat polymorphism D21S1245 demonstrates hypermutability in germline and somatic cells. Hum Mol Genet 4:1193–1199

Underhill PA, Jin L, Zemans R, Oefner PL, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implication for human evolutionary history. Proc Natl Acad Sci USA 93:196–200

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellites loci: the stepwise mutation model revised. Genetics 133:737–749

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769–779

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, et al (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183

Zhang L, Leeflang EP, Yu J, Arnheim N (1994) Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. Nat Genet 7:531–535