

RMplex reveals population differences in RM Y-STR mutation rates and provides improved father-son differentiation in Japanese

Tomomi Otagiri, Noriko Sato, Hideki Asamura, Evelina Parvanova, Manfred Kayser, Arwin Ralf



PII: S1872-4973(22)00107-7

DOI: <https://doi.org/10.1016/j.fsigen.2022.102766>

Reference: FSIGEN102766

To appear in: *Forensic Science International: Genetics*

Received date: 11 July 2022

Accepted date: 15 August 2022

Please cite this article as: Tomomi Otagiri, Noriko Sato, Hideki Asamura, Evelina Parvanova, Manfred Kayser and Arwin Ralf, RMplex reveals population differences in RM Y-STR mutation rates and provides improved father-son differentiation in Japanese, *Forensic Science International: Genetics*, (2022) doi:<https://doi.org/10.1016/j.fsigen.2022.102766>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier.

RMplex reveals population differences in RM Y-STR mutation rates and provides improved father-son differentiation in Japanese

Tomomi Otagiri¹, Noriko Sato¹, Hideki Asamura¹, Evelina Parvanova², Manfred Kayser², Arwin Ralf^{2*}

¹ Department of Legal Medicine, Shinshu University School of Medicine, Matsumoto, Nagano, Japan

² Department of Genetic Identification, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands

*Corresponding author: a.ralf@erasmusmc.nl

Abstract

Rapidly mutating Y chromosomal short tandem repeat markers (RM Y-STRs) –characterized by at least one mutation per 100 generations– are suitable for differentiating both related and unrelated males. The recently introduced multiplex method RMplex allows for the efficient analysis of 30 Y-STRs with increased mutation rates, including all 26 currently known RM Y-STRs. While currently available RM Y-STR mutation rates were established mostly from European individuals, here we applied RMplex to DNA samples of 178 genetically confirmed father-son pairs from East Asia. For several Y-STRs, we found significantly higher mutation rates in Japanese compared to previous estimates. The consequent father-son differentiation rate based on RMplex was significantly higher (52%) in Japanese than previously reported for Europeans (42%), and much higher than with Yfiler Plus in both sample sets (14% and 13%, respectively). Further analysis suggests that the higher mutation and relative differentiation rates in Japanese can in part be explained by on average longer Y-STR alleles relative to Europeans. Moreover, we show that the most striking difference, which was found in DYS712, could be linked to a Y-SNP haplogroup (O1b2-P49) that is common in Japanese and rare in other populations. We encourage the forensic Y-STR community to generate more RMplex data from more population samples of sufficiently large sample size in combination with Y-SNP data to further investigate population effects on mutation and relative differentiation rates. Until more RMplex data from more populations become available,

caution shall be placed when applying RM Y-STR mutation rate estimates established in one population, such as Europeans, to forensic casework involving male suspects of paternal origin from other populations, such as non-Europeans.

Keywords:

Y-STR; Rapidly mutating Y-STRs; RM Y-STRs; RMplex; Mutation rate; Male relative differentiation; Population differences

1. Introduction

Y chromosomal short tandem repeats (Y-STRs) are commonly used in forensic DNA analysis, especially in sexual assault cases involving mixed DNA evidence to which the male perpetrator and the female victim contributed, as such male-female mixed DNA samples are notorious for showing difficulties in perpetrator identification based on autosomal STR profiling [1]. However, due to the paternal inheritance of the male-specific part of the Y chromosome, Y-STR profiles are typically shared between patrilineally related men. In consequence, a Y-STR haplotype match between the male suspect and the evidence DNA does not necessarily indicate that the suspect contributed to the evidence sample, as this could also have been his male relatives that share the same Y-STR haplotype. The need to overcome this limitation motivated the search for Y-STRs characterized by elevated mutation rates [2], 13 Y-STRs which had mutation rates in the order of 10^{-2} (at least 1 mutation every 100 generations per locus, mpg) were identified and termed rapidly mutating Y-STRs (RM Y-STRs) [3]. Recently, a new set of RM Y-STRs was discovered by applying *in silico* search for candidate markers and empirical confirmation in father-son pairs [4], which increased the number of currently known RM-Y-STRs to 26 and also identified Y-STRs with mutation rates higher than 5×10^{-3} but lower than 10^{-2} termed fast mutating Y-STRs (FM Y-STRs). Moreover, a new multiplex genotyping assay, RMplex, was developed and validated that analyses a total of 30 Y-STRs with increased mutation rates, which includes the 26 RM Y-STRs and 4 FM Y-STRs [5]. It was shown that this new RMplex assay outperformed the state-of-the art commercial Y-STR genotyping

assay Yfiler™ Plus PCR Amplification Kit, demonstrating increased father-son differentiation rate by a factor of three [6].

Notably, all currently known RM Y-STRs were discovered solely via mutation rate studies in Europeans [2, 4]. While some worldwide father-son and population data were previously established for the initial set of 13 RM Y-STRs [7], non-European data for the new set of 26 RM Y-STRs and for the full set of 30 Y-STRs included in RMplex are missing as of yet. Some previous studies showed differences in mutation rates between males from populations for the initial set of 13 RM Y-STRs [7-13]. While most previous Y-STR mutation rate studies lack Y-SNP haplogroup data, Claerhout *et al.* studied the effect of Y-SNP haplogroups on Y-STR mutation rates in European males [14] and found lower overall average mutation rates within some Y-SNP haplogroups (i.e., haplogroup I & J) compared to others (i.e., R1b). Lower Y-STR mutation rates coincided with allele frequency distributions that were skewed towards shorter alleles for some Y-STRs in males with haplogroup I & J. Other previous STR mutation studies also revealed a strong impact of allele length, i.e., number of repeats, on mutability of STRs, with longer alleles leading to increased and shorter ones to decreased mutation rates [15, 16]. Many populations have varying Y-SNP haplogroup compositions indicating different male founders who carried their own Y-STR haplotypes. Male founders with longer Y-STR alleles may lead to increased mutability of such Y-STRs in the population of descendants, but data evidence to empirically prove this hypothesis is scarce.

Y-STR mutations are rare events and thus mutation rate estimates are prone to stochastic effects. Hence, ideally, thousands of father-son pairs should be analyzed to make accurate Y-STR mutation rate estimates in samples from many populations; in reality, however, most studies only involve hundreds. Obviously, the larger the mutation rate of a Y-STR, the more mutations will be observed, thereby decreasing the impact of stochastic effects. This makes RM Y-STRs the most suitable markers for studying population effects on Y-STR mutation and male relative differentiation rates, even when sample size is limited.

A recent study applied the current state-of-the-art commercial Yfiler™ Plus PCR Amplification Kit (Yfiler) to Japanese father-son pairs and found that the six RM Y-STR markers included in this commercial kit did not show the expected elevated mutation rates known from previous European studies [17]. The authors therefore concluded that this commercial kit may be less suitable for the purpose of male relative differentiation in the Japanese population.

In the current study, we applied RMplex to DNA-confirmed father-son pairs originating from Japan. The newly established mutation and relative differentiation rate data were compared to previously published consensus mutation rate estimates based on multiple studies [6]. Additionally, Y haplogroup data were generated from genotyping 10 Y-SNPs to infer 8 Y haplogroups known to be frequent in Japanese, which allowed us to link the established Y-STR mutation data with the relevant Y-SNP haplogroup background information.

2. Materials and methods

2.1 DNA samples

Genomic DNA was extracted of a total of 340 males using QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's instructions. Of the 340 males were included in this study, 296 originated from pairs of fathers with a single son (n=148), while additionally the dataset consisted of one father with three sons, and 12 fathers with two sons. Lastly, the dataset included one family with four generations of male relatives (one individual per generation, three father-son pairs). In total, 178 father-son pairs were defined from this dataset. The biological relationship of all pair of males was confirmed through autosomal STR typing. This study was approved by the Ethics committee of Shinshu University School of Medicine (Permission number: 667). All individuals included in this study provided informed consent. A total of 280 samples used in the present study overlap with those in the previous study based on Yfiler Plus [17], while the remaining 60 samples were newly typed in the context of the present study.

2.2 *Amplification and genotyping*

Y-STR typing was done using the protocol previously described by Ralf et al. [5], while using the alternative forward primer for DYS570 as described in that publication. The total reaction volume was reduced to 10 μ L, which keeping the primer concentrations the same as described previously [5]. Amplification was performed using a GeneAmp[®] PCR System 9700 (Thermo Fisher Scientific), the resulting PCR products were analyzed using a 3500 Genetic Analyzer (Thermo Fisher Scientific) using POP-4 Polymer (Thermo Fisher Scientific) and the resulting electropherograms were analyzed using GeneMapper[®] ID-X Software v1.4 (Thermo Fisher Scientific).

2.3 *Y-SNP based haplogroup analysis*

Y-SNP testing was performed by developing a custom-made multiplex genotyping assay using SNaPshot[™] Multiplex Kit (Thermo Fisher Scientific). A total of 10 Y-SNPs were selected, 8 target specific (sub)haplogroups that were expected to be present in the Japanese male population: C-M130, D-M174, N-M231, O1a-M119, O1b2-P49, O2-M122, O-P186*(xM119,P49,M122), Q-M242). Additional two intermediate Y-SNPs were included in the design: DE-M145 and CF-P143. The primer sequences and thermal cycler conditions are shown in supplementary Table S1. The PCR was performed in 10 μ L volumes including 5 μ L QIAGEN Multiplex PCR Plus Kit (Qiagen), PCR primers with the concentrations as detailed in supplementary Table S1 and 1 μ L of DNA (\sim 1 ng/ μ L). The PCR products were purified by adding 4 μ L of 1 unit / μ L EXO-SAP (Thermo Fisher Scientific) to the total volume of PCR products and incubating for 30 min at 37°C, followed by 15 min at 80°C for deactivation. SBE reactions were performed as described in Table S1 in a mix consisting of 2 μ L SNaPshot Ready Reaction Mix (Thermo Fisher Scientific), 2 μ L of primer mix, 5 μ L of DNA-free water and 1 μ L of purified PCR products. The SBE products were purified by adding 2 μ L of 1 unit / μ L SAP (Thermo Fisher Scientific) to the total volume of

SBE products and incubating for 30 min at 37°C, followed by 15 min at 80°C for deactivation.

Amplifications, enzymatic purifications and single base pair extensions were performed using a GeneAmp® PCR System 9700 (Thermo Fisher Scientific), the resulting PCR products were analyzed using a 3500 Genetic Analyzer (Thermo Fisher Scientific) using POP-4 Polymer (Thermo Fisher Scientific) and the resulting electropherograms were analyzed using GeneMapper® ID-X Software v1.4 (Thermo Fisher Scientific).

2.4 *Data analysis*

Both the mutation rates and the differentiation rates were calculated using the frequentist approach.

The Clopper-Pearson interval was used to determine the 95% confidence intervals of these rates.

Fisher's exact tests were used to determine the statistical significance of difference observed between the present study and previous studies. Pairwise Rst values were calculated by an in-house pipeline that performs per-marker allele comparisons for each pair of samples. The Rst value for a given pair was defined as the sum of the differences among all Y-STRs. Multi-copy markers pose additional complexity for this approach because it is typically not possible to tell which copies correspond to each other. To calculate the Rst value in such Y-STRs the pipeline chose the shortest path, e.g. if one individual typed with alleles 12, 16 was paired with another individual displaying alleles 13, 15, the pipeline derived a distance value of 2 (the sum of the difference between 12 and 13, and the difference between 15 and 16) instead of 6 (the sum of the difference between 12 and 15 and the difference between 13 and 16). R [18] was used to create boxplots.

3. **Results and discussion**

3.1 *Mutation analysis*

In total, 157 mutations were observed amongst the 178 Japanese father-son pairs, of which 138 were detected using RMplex and 29 with Yfiler Plus of which 10 were found at the six RM Y-STRs overlapping between the two methods. All mutations with the specific allele changes and the haplogroup of the pairs in which they occurred are shown in supplementary Table S2. A total of two 2-step mutations were observed, one in DYF399S1 and one in DYS712, the remaining 155 mutations (98.7%) were single step mutations. Generally speaking, the sample size in the present study was, with 178 father-son pairs, relatively small to yield highly reliable estimations of the mutation rates. This uncertainty becomes apparent from the 95% confidence intervals where even Y-STRs not showing a single mutation among the 178 pairs have an upper limit of 2×10^{-2} mpg (Table 1). Because of this limited sample size, all conclusions ought to be treated with caution, at least until larger scale studies in the same population replicate these results.

The Y-STR marker that showed most mutations was DYF399S1 with 19 mutations, which confirms previous studies in different populations showing that this multi-copy RM Y-STR is the most mutable out of all previously characterized Y-STR [2, 4, 6-13]. However, in the current study DYF399S1 mutated even more frequently than previously observed with >10% of the analyzed pairs showing a mutation at this RM Y-STR (Table 1). The noted increase in mutation rate was statistically significant (p-value: 0.028) relative to the reference mutation rate based on >7,500 father-son pairs of which only 6.3% displayed a mutation for this multi-copy marker [6]. Another remarkable result was the high number of 17 mutations found at DYS712, which resulted in a mutation rate estimate of 9.6×10^{-2} mpg. This mutation rate we obtained for DYS712 here from Japanese father-son pair data is significantly higher than those previously obtained from European father-son pair data with 2.7×10^{-2} mpg (p-value: <0.0001) [4] and 4.3×10^{-2} mpg (p-value: 0.0138) [6]. A study analyzing father-son pairs from the Shanxi Province in China reported a mutation rate of 3.0×10^{-2} mpg for DYS712 [19], which is similar to the

rates previously obtained from European data for this marker, but significantly lower than the rate we obtained here for Japanese (p-value: 0.0030).

Notably, the commonly used Y-STR DYS458 showed a mutation rate of 3.4×10^{-2} mpg in the present study, which is remarkably higher compared to the rate of 8.4×10^{-3} mpg that was previously estimated in a study using European father-son pairs [2]. Also, compared to the consensus estimate based on data from 11,830 father son-pairs [6], the observed mutation rate in the current Japanese study is significantly higher (p-value: 0.0051) than the consensus mutation rate of 8.5×10^{-3} that was previously obtained (Table 1). Other Y-STRs that show significant differences compared to the previously reported consensus mutation rate estimates [6] are: DYF1001 (p-value 0.0120), DYS460 (p-value 0.0092), and DYS713 (p-value 0.0162). All six Y-STRs showed an increased mutation rate in the present study compared to the previous consensus estimates. Table 1 shows mutation rate estimates as obtained in the present study; these mutation rates are compared to the consensus Y-STR mutation rate estimates as described in Neuhuber et al., 2022 [6].

Table 1: Empirically established locus-specific mutation rates of 49 Y-STRs by applying RMplex and Yfiler™ Plus to a total of 178 DNA-confirmed father-son pairs from Japan and their comparisons with consensus locus-specific reference mutation rates previously described [6].

Marker	Assay	Total pairs	Mutations	Expansions	Contractions	Mutation rate (x10 ⁻³)	95% confidence interval (x10 ⁻³)	Reference mutation rate (x10 ⁻³)	p-value #
DYF399							65.5-		0.028
S1	RMplex	178	19	7	12	106.7	161.7	62.8	0
DYS712	RMplex	178	17	8	9	95.5	148.5	31.1	1
DYF100							56.6-		0.012
1	RMplex	178	17	11	6	95.5	148.5	48.0	0
DYF403							23.4-		0.098
S1a	RMplex	178	9	5	4	50.6	93.8	27.3	5
DYF100							19.6-		0.530
0	RMplex	178	8	4	4	44.9	86.6	35.9	5
DYS713	RMplex	178	7	1	6	39.3	16-79.3	13.9	2
DYS458	Yfiler Plus	178	6	3	3	33.7	12.5-71.9	8.5	1
DYS724	RMplex	178	6	2	4	33.7	12.5-71.9	48.0	3
DYS101							9.2-		0.185
0	RMplex	178	5	1	4	28.1	64.3	14.0	2
DYS711	RMplex	178	4	2	2	22.5	6.2-56.5	26.6	0
DYS612	RMplex	178	4	1	3	22.5	6.2-56.5	16.3	1
DYF403							6.2-		0.085
S1b	RMplex	178	4	4	0	22.5	56.5	9.1	9
DYS100							6.2-		0.119
5	RMplex	178	4	0	4	22.5	56.5	9.8	1
DYS576	Yfiler Plus+RMplex	178	4	1	3	22.5	6.2-56.5	12.7	8
DYS460	Yfiler Plus	178	4	2	2	22.5	6.2-56.5	4.3	2
DYS547	RMplex	178	3	3	0	16.9	3.5-48.5	14.7	4
DYS100							3.5-		1.000
7	RMplex	178	3	2	1	16.9	48.5	17.2	0
DYF404							3.5-		0.492
S1	RMplex	178	3	1	2	16.9	48.5	12.5	1
DYS101									1.000
3	RMplex	178	2	1	1	11.2	1.4-40	10.8	0

DYR88	RMplex	178	2	1	1	11.2	40.0	26.3	0	1.4-	0.317
DYS526										1.4-	1.000
b	RMplex	178	2	1	1	11.2	40.0	12.3	0	1.4-	0.765
DYF100										1.4-	0.661
2	RMplex	178	2	1	1	11.2	40.0	16.8	0	1.4-	0.661
	Yfiler									1.4-	0.661
DYS570	Plus+RMplex	178	2	2	0	11.2	40.0	8.3	2	1.4-	1.000
DYS100										1.4-	1.000
3	RMplex	178	2	1	1	11.2	40.0	12.6	0	1.4-	0.214
DYS481	Yfiler Plus	178	2	1	1	11.2	40.0	4.7	1	1.4-	1.000
DYS101										1.4-	1.000
2	RMplex	178	2	1	1	11.2	40.0	15.8	0	1.4-	0.667
DYS626	RMplex	178	2	0	2	11.2	40.0	8.6	6	1.4-	1.000
	Yfiler									1.4-	1.000
DYS627	Plus+RMplex	178	2	2	0	11.2	40.0	14.5	0	0.1-	0.725
	Yfiler									0.1-	0.725
DYS449	Plus+RMplex	178	1	0	1	5.6	30.9	11.2	8	0.1-	1.000
DYS385	Yfiler Plus	178	1	1	0	5.6	30.9	7.5	0	0.1-	1.000
DYF387	Yfiler									0.1-	1.000
S1	Plus+RMplex	178	1	0	1	5.6	30.9	10.2	0	0.1-	0.544
DYS456	Yfiler Plus	178	1	1	0	5.6	30.9	4.4	3	0.1-	0.271
DYS393	Yfiler Plus	178	1	1	0	5.6	30.9	1.7	9	0.1-	0.304
DYS19	Yfiler Plus	178	1	0	1	5.6	30.9	2.0	1	0.1-	0.298
YGATA										0.1-	0.298
H4	Yfiler Plus	178	1	1	0	5.6	30.9	1.9	0	0.1-	1.000
DYS442	RMplex	178	1	0	1	5.6	30.9	7.4	0	0.1-	0.500
DYS635	Yfiler Plus	178	1	1	0	5.6	30.9	3.8	5	0.1-	0.626
DYS389										0.1-	0.626
II	Yfiler Plus	178	1	0	1	5.6	30.9	5.5	7		1.000
DYS533	Yfiler Plus	178	0	0	0	0.0	0-20.5	3.5	0		1.000
DYS439	Yfiler Plus	178	0	0	0	0.0	0-20.5	4.8	0		1.000
DYS391	Yfiler Plus	178	0	0	0	0.0	0-20.5	2.5	0		0.178
	Yfiler										0.178
DYS518	Plus+RMplex	178	0	0	0	0.0	0-20.5	13.3	4		1.000
DYS437	Yfiler Plus	178	0	0	0	0.0	0-20.5	1.2	0		1.000

DYF393									0.624
S1	RMplex	178	0	0	0	0.0	0-20.5	7.1	8
DYS389									1.000
I	Yfiler Plus	178	0	0	0	0.0	0-20.5	2.4	0
									1.000
DYS448	Yfiler Plus	178	0	0	0	0.0	0-20.5	0.8	0
									1.000
DYS390	Yfiler Plus	178	0	0	0	0.0	0-20.5	2.7	0
									1.000
DYS438	Yfiler Plus	178	0	0	0	0.0	0-20.5	0.3	0
									1.000
DYS392	Yfiler Plus	178	0	0	0	0.0	0-20.5	0.8	0

statistically significant differences (p-values <0.05) are indicated in bold

3.2 Differentiation of father-son pairs

Based on Yfiler Plus alone, a total of 25 out of the 178 (14%) Japanese father-son pairs were differentiated, which is comparable to a previous study based on European males where 13% of father-son pairs were differentiated [6]. Based on RMplex alone, a total of 93 out of the 178 pairs (52%) were differentiated, which reflects an about 3.7-fold increase compared to Yfiler Plus in the same samples, and is significantly higher (Fisher's exact p-value: 0.0179) than in a previous study based on European father-son pairs where a differentiation rate of 42% was reported [6]. Combining the data from both Yfiler Plus and RMplex resulted in an even higher father-son differentiation rate of 57% with 101 out of the 178 Japanese pairs being separated. With less than a five percent point increase, the contribution of the non-overlapping Yfiler Plus markers to the differentiation of male relatives was limited, as expected based on their lower mutation rates.

When looking more closely to the number of mutations that differentiate a father-son pair, we saw that 22 of the 25 pairs differentiated by Yfiler Plus (88%) only showed a mutation at a single Y-STR marker. In contrast, for RMplex and for both assays combined, 62% of the 93 and 57% of the 101 differentiated pairs were separated only by a single mutation, respectively. Furthermore, 8%, 27%, and

33% of the differentiated pairs were separated by mutations at two Y-STRs and 4%, 11%, and 8% of the differentiated pairs by mutations at three Y-STRs, for Yfiler Plus, RMplex, and the combined methods, respectively (Figure 1). Mutations at four and five markers were only observed when both methods were combined and were each observed only in a single pair.

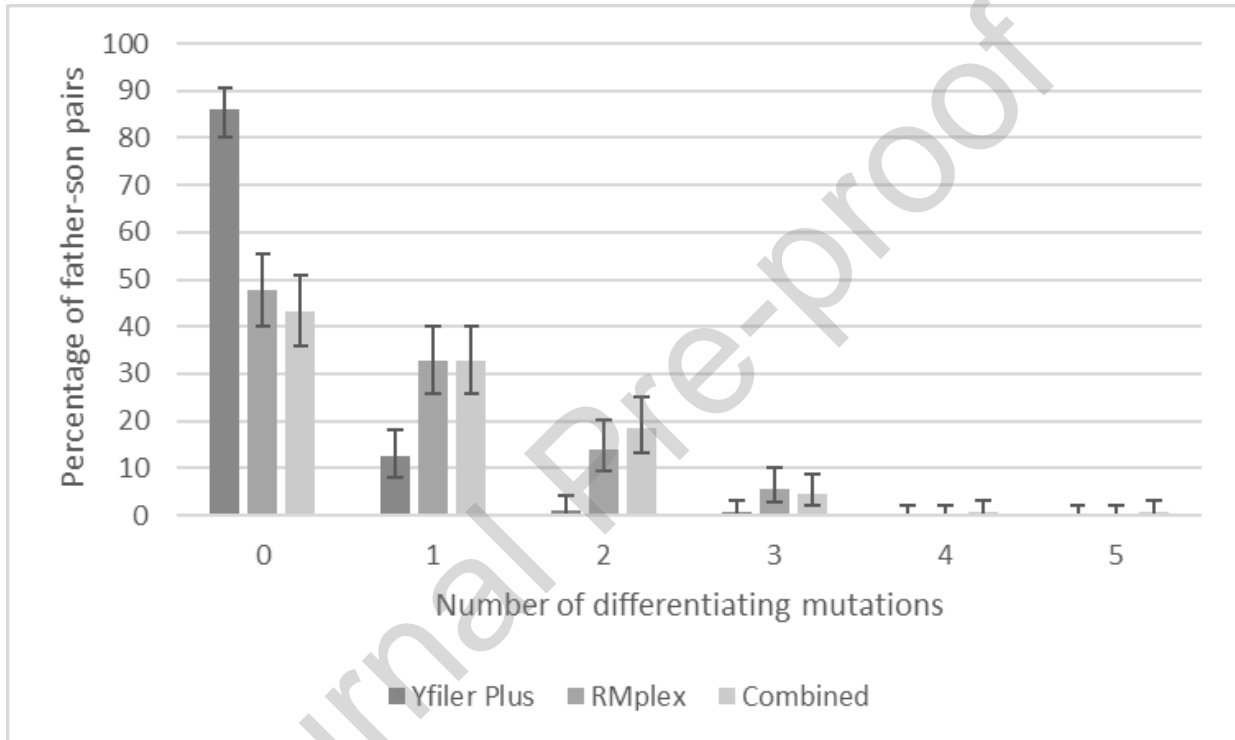


Figure 1. Percentage of father-son pairs analyzed with Yfiler Plus (25 Y-STRs), RMplex (30 Y-STRs), and both methods combined (49 Y-STRs) with mutations at zero, one, two, three, four, and five Y-STR markers per pair. None of these pairs was differentiated by mutations at more than five Y-STRs. The error bars represent the exact binomial 95% confidence interval (Clopper-Pearson).

3.3 Differentiation of unrelated males

To assess the efficiency in differentiated unrelated males, we compared the Y-STR haplotypes obtained with Yfiler Plus and RMplex combined in the total of 162 fathers and found that each of them carried a unique haplotype based on the full set of 49 Y-STR. Notably, the same number of unique haplotypes were also seen when considering RMplex and Yfiler Plus separately. Hence, in the current study, no

difference in capabilities to differentiate unrelated males was seen for RMplex and Yfiler Plus, which was also reported in a previous European study [6]. However, the relatively low sample size in both studies might have influenced this rather unexpected result, as the probability of observing shared haplotypes too increases with sample size. Future studies with increased sample size need to show if indeed both methods are performing equally well in differentiating unrelated men, or if the identical performance of both methods was influenced by sample size effects in this Japanese and the previous European study.

Despite its limited sample size, the data obtained in the current study does provide some insights regarding the potential to differentiate unrelated males by determining pairwise R_{st} values. R_{st} considers the mutational differences between haplotypes and is typically estimated from data of different population to express the proportion of the diversity seen between populations. Here we estimated R_{st} between pairs of individual haplotypes, and not between populations as is typically done. This way, R_{st} provides an estimate on the diversity difference between the haplotypes derived from the two genotyping methods. As shown in Figure 2, there is a sharp distinction between Yfiler Plus and RMplex in pairwise- R_{st} distributions, where Yfiler Plus clearly results in lower R_{st} values, indicating more similarity between the Yfiler Plus derived haplotypes from the unrelated males compared to those from RMplex. By extrapolation, it could be expected that more similarity in a small sample would translate to more overlapping haplotypes in a significantly larger sample. Although, there is still a need for empirical evidence based on large numbers of unrelated males, the difference in R_{st} values between the two methods could be seen as a first indication that RMplex may be superior in differentiating not only related males but also unrelated ones.

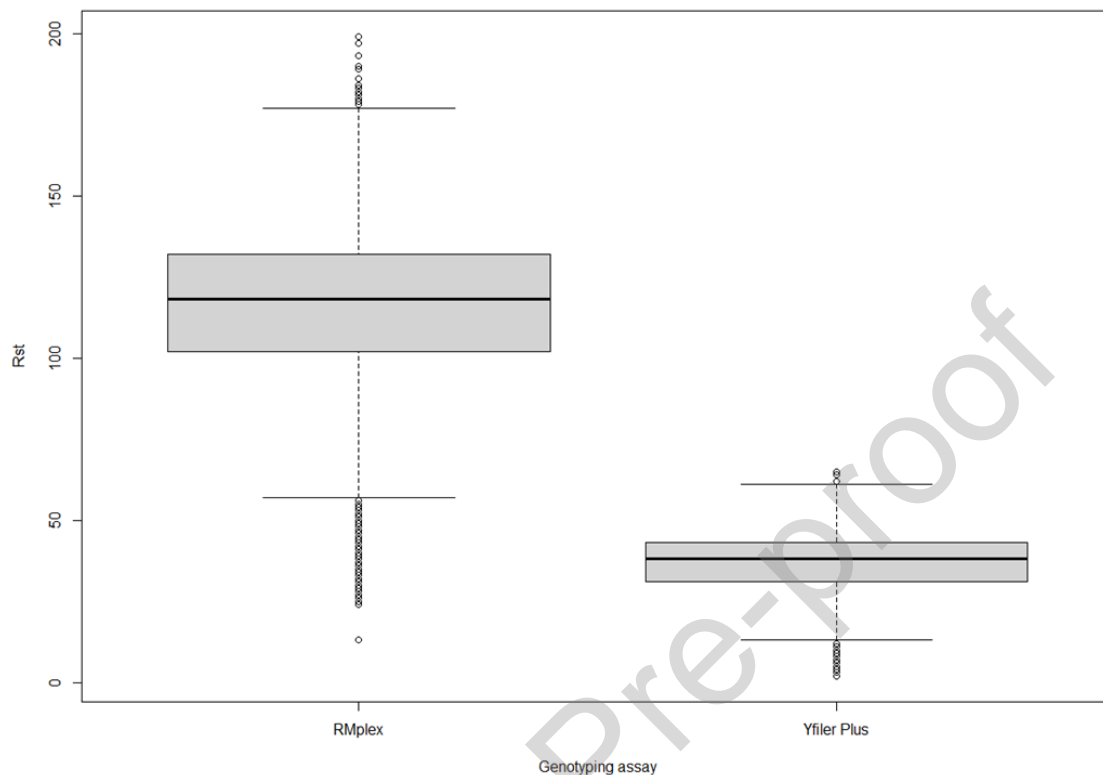


Figure 2. Pairwise Rst-value distribution obtained RMplex and Yfiler Plus based on the 162 unrelated Japanese males.

3.4 Differences in Y-STR allele lengths between populations and Y-SNP haplogroups

Y-SNP analysis was performed on 162 unrelated fathers with one individual not providing a full profile, leaving 161 unrelated males in this analysis for which Y-SNP based haplogroups were established. The most commonly observed haplogroup in this Japanese dataset was O1b2-P49 (32%), followed by D-M174 (30%), O2-M122 (19%), C-M130 (14%). The remaining four haplogroups were found in less than 5 individuals each (<2.5%). A previous study that analyzed Y-SNPs in commonly observed Japanese surnames found that 37% belonged to haplogroup D, 30% to O1b, 20% to O2, and 9% to C [20]. Despite some stochastic variations, these results show a similar occurrences of the most common haplogroups in the Japanese population.

One of the most remarkable results obtained in the present study was the increased number of mutations found at DYS712. We compared the allele frequencies previously reported for DYS712 from a European population sample from Austria [6], to the allele frequencies found in the current Japanese samples. Figure 3 shows that there was more variability in the DYS712 Y-STR alleles found in the Japanese relative to the Europeans. Moreover, the Japanese sample had a higher median allele length compared to the European. When comparing the most common Y-SNP haplogroups in the Japanese sample set (Figure 3), it became evident that the longer Y-STR alleles at DYS712 were especially common in males belonging to a subgroup of haplogroup O1b2 (O-P49). This was the most frequently observed haplogroup in the current Japanese sample and is completely absent from Europeans [21]. It is widely established that longer STR alleles are more prone to mutations [15, 16]. This notion is supported by the fact that in the current study 13 out of the total 17 mutations (76%) observed at DYS712 arose from fathers that had an allele length larger than the population median value of 23 in this Japanese dataset. That this effect likely is haplogroup dependent is further supported by our finding that 9 out of 17 pairs (53%) that showed a mutation for DYS712 belonged to haplogroup O1b2 (O-P49). Notably, this haplogroup was rarely found in Eastern Han Chinese [22], which may explain why the high mutation rate found in the current Japanese dataset for DYS712 was not previously found in Chinese from the Shanxi region [23]. Also the opposite effect is seen in the Japanese data, where the frequently occurring haplogroup D generally displays short Y-STR alleles at DYS712 (Figure 3), while only 1 out of the 17 pairs (6%) that showed a mutation at DYS712 belonged to haplogroup D despite the high overall prevalence of haplogroup D in the studied population sample. Allele frequencies for all 49 Y-STRs analyzed with RMplex and Yfiler Plus are shown in supplementary Table S3; additionally the haplotype frequencies of the multi-copy are shown in Table S4.

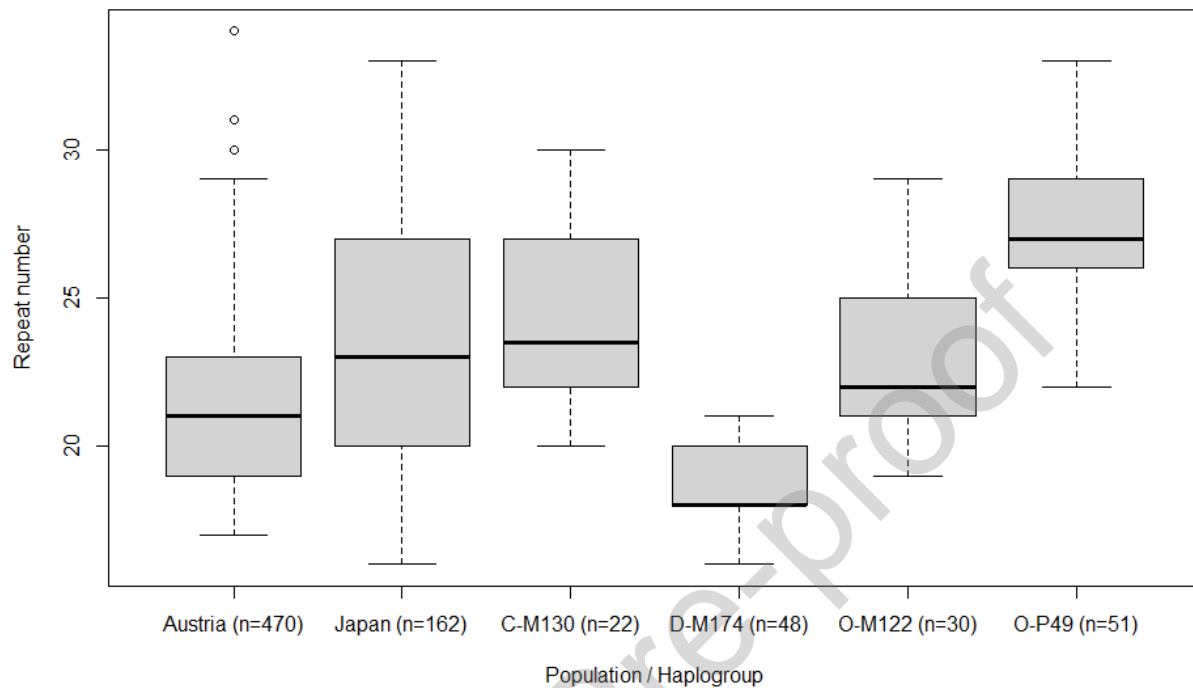


Figure 3. Boxplots showing the difference in allele length distribution of the RM Y-STR marker DYS712 between the current Japanese study (overall and stratified per predicted Y-SNP haplogroup) where this marker has a mutation rate of 9.6×10^{-2} and a previous European study where the mutation rate was 3.1×10^{-2} [6].

To further investigate the potential influence of allele frequencies on mutation rates, we compared both the mutation rates and the mean allele frequency from both the current Japanese study and the previous European study that used RMplex [6] for the six Y-STRs that showed significantly different mutation rates between the current study and the previously established consensus mutation rates (Table 1). As evident from Table 2, the observed mutation rate differences were not always statistically significant in the direct comparison. Both studies had a relatively small sample size leading to reduced statistical power, whereas the sample size of the consensus mutation rate estimates used in prior comparisons were relatively large. Regardless, the higher occurrence of mutations in the current study on Japanese father-son pairs compared to the study based on European pairs was still clearly noticeable

(Table 2). Interestingly, for all six Y-STRs, the mean allele length was also higher in the Japanese population than it was in the European population, albeit to different extents across the six markers (Table 2).

Table 2: Direct comparison of the mutation rates and mean allele lengths between the current Japanese study and a previous European study [6] for six Y-STRs with remarkable high mutation rate estimates in the current Japanese study.

Marker	Mutation rates ($\times 10^{-3}$)			Mean allele length	
	Japanese	European	p-value [#]	Japanese	European
DYF399S1	106.7	77.4	0.217	23.9	23.1
DYF1001	95.5	36.0	0.005	64.3	62.5
DYS712	95.5	43.4	0.014	23.4	21.4
DYS713	39.3	17.0	0.139	46.1	43.0
DYS458	33.7	13.2	0.103	17.0	16.2
DYS460	22.5	7.5	0.115	10.7	10.5

[#] The p-values are based on Fisher's exact tests and the proportions of the father-son pairs where mutations were observed from each study, respectively.

Due to limited sample size, the evidence we present here for increased RM Y-STR mutation and relative differentiation rates in Japanese compared to Europeans being explained by population effects is statistically not clear-cut. However, our findings show a clear trend, which shall be confirmed with more Japanese data in future studies. Moreover, also other population samples of paternally related males from different parts of the world and of suitable size should be analyzed with RMplex and relevant Y-SNPs to collect more empirical evidence on whether population differences in RM Y-STR mutation and relative differentiation rates truly exist and can be explained.

4. Conclusion

Here, we showed for the first time, the efficiency at which RMplex differentiates non-European, i.e., Japanese, father-son pairs, which turned out to be at a significantly higher rate (52%) than previously established in Europeans (42%), and much higher than with the current state-of-the-art commercial Y-

STR kit Yfiler Plus (14%). Furthermore, we show that Y-STR mutation rates may depend on Y-STR allele lengths and that this effect can be linked to Y-SNP haplogroup background. By showing preliminary evidence for population differences in RM Y-STR mutation and male relative differentiation rates, our study highlights that the rates obtained from one population may not necessarily be transferable to samples from another population. For forensic practice this means that using RM Y-STR mutation rates established in for instance Europeans for interpreting RMplex casework results from a suspect of non-European paternal ancestry may be error-prone. Additionally, our study calls for more RMplex population studies of suitably large (i.e., larger than used here) sample size to further study potential population effects on RM Y-STR mutation and relative differentiation rates, where Y-SNP haplogroups should be analyzed too to better understand the rate differences that may be observed between different populations. Ultimately, our study implies that there may be a need to look beyond locus-specific mutation rates and rather start establishing population-specific, haplogroup-specific, or perhaps allele-specific mutation rates for each locus.

References

1. Kayser, M., *Forensic use of Y-chromosome DNA: a general overview*. Human Genetics, 2017. **136**(5): p. 621-635.
2. Ballantyne, K.N., et al., *Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications*. The American Journal of Human Genetics, 2010. **87**(3): p. 341-353.
3. Ballantyne, K.N., et al., *A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages*. Forensic Science International: Genetics, 2012. **6**(2): p. 208-218.
4. Ralf, A., et al., *Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers*. Human Mutation, 2020. **41**(9): p. 1680-1696.
5. Ralf, A., et al., *RMplex: An efficient method for analyzing 30 Y-STRs with high mutation rates*. Forensic Science International: Genetics, 2021(55): p. 102595.
6. Neuhuber, F., et al., *Improving the differentiation of closely related males by RMplex analysis of 30 Y-STRs with high mutation rates*. Forensic Science International: Genetics, 2022: p. 102682.
7. Ballantyne, K.N., et al., *Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats*. Human Mutation, 2014. **35**(8): p. 1021-1032.

8. Adnan, A., et al., *Improving empirical evidence on differentiating closely related men with RM Y-STRs: a comprehensive pedigree study from Pakistan*. Forensic Science International: Genetics, 2016. **25**: p. 45-51.
9. Boattini, A., et al., *Mutation rates and discriminating power for 13 rapidly-mutating Y-STRs between related and unrelated individuals*. PLOS One, 2016. **11**(11): p. e0165678.
10. Chen, Y., et al., *Mutation rates of 13 RM Y-STRs in a Han population from Shandong province, China*. Forensic Science International: Genetics Supplement Series, 2017. **6**: p. e346-e348.
11. Yuan, L., et al., *Mutation analysis of 13 RM Y-STR loci in Han population from Beijing of China*. International Journal of Legal Medicine, 2019. **133**(1): p. 59-63.
12. Zgonjanin, D., et al., *Mutation rate at 13 rapidly mutating Y-STR loci in the population of Serbia*. Forensic Science International: Genetics Supplement Series, 2017. **6**: p. e377-e379.
13. Zhang, W., et al., *Multiplex assay development and mutation rate analysis for 13 RM Y-STRs in Chinese Han population*. International Journal of Legal Medicine, 2017. **131**(2): p. 345-350.
14. Claerhout, S., et al., *Determining Y-STR mutation rates in deep-routing genealogies: identification of haplogroup differences*. Forensic Science International: Genetics, 2018.
15. Kelkar, Y.D., et al., *The genome-wide determinants of human and chimpanzee microsatellite evolution*. Genome Research, 2008. **18**(1): p. 30-38.
16. Brinkmann, B., et al., *Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat*. The American Journal of Human Genetics, 1998. **62**(6): p. 1408-1415.
17. Otagiri, T., et al., *Mutation analysis for 25 Y-STR markers in Japanese population*. Legal Medicine, 2021. **50**: p. 101860.
18. R Core Team, *R: A language and environment for statistical computing*. 2013.
19. Liu, J., et al., *The construction and application of a new 17-plex Y-STR system using universal fluorescent PCR*. International Journal of Legal Medicine, 2020. **134**(6): p. 2015-2027.
20. Ochiai, E., et al., *Y chromosome analysis for common surnames in the Japanese male population*. Journal of Human Genetics, 2021. **66**(7): p. 731-738.
21. Navarro-López, B., et al., *Phylogeographic review of Y chromosome haplogroups in Europe*. International Journal of Legal Medicine, 2021. **135**(5): p. 1675-1684.
22. Lang, M., et al., *Forensic characteristics and genetic analysis of both 27 Y-STRs and 143 Y-SNPs in Eastern Han Chinese population*. Forensic Science International: Genetics, 2019. **42**: p. e13-e20.
23. Liu, J., et al., *Development of a new 17 Y-STRs system using fluorescent-labelled universal primers and its application in Shanxi population in China*. Forensic Science International: Genetics Supplement Series, 2019. **7**(1): p. 95-97.

Highlights

- RMplex differentiated more father-son pairs in Japanese than in Europeans
- RMplex differentiated many more father-son pairs than Yfiler Plus
- Strong mutation rate differences were found compared to previous studies
- Mutation rate population differences may depend on Y-SNP haplogroup background

Highlights

- RMplex differentiated more father-son pairs in Japanese than in Europeans
- RMplex differentiated many more father-son pairs than Yfiler Plus
- Strong mutation rate differences were found compared to previous studies
- Mutation rate population differences may depend on Y-SNP haplogroup background

Journal Pre-proof