

ARTICLE

Apparent intrachromosomal exchange on the human Y chromosome explained by population history

Ralf Kittler¹, Axel Erler¹, Silke Brauer¹, Mark Stoneking¹ and Manfred Kayser^{*,1}

¹Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany

The human Y chromosome displays an unusual content of repetitive sequences. Y-chromosomal repeats are potential targets for intrachromosomal recombination, which is thought to be involved in a number of Y-associated defects, such as male infertility. Such rearrangements could potentially be investigated by the use of highly polymorphic DNA markers located within the repeat units, such as microsatellites. Here we analyse the two copies of the Y-chromosomal microsatellite DYS385, which we identified and localized to an ~190 kb duplicated and inverted fragment at Yq11.223. We found a highly significant correlation ($r=0.853$, $P<0.001$) and a nonsignificant difference in a χ^2 -test ($\chi^2=15.45$, $P>0.05$) between the allele frequency distributions at both copies of the Y-STR in a German population sample ($n=70$). Such nearly identical allele frequency distribution between two copies of a duplicated highly polymorphic microsatellite cannot be explained by the independent mutational process that creates microsatellite alleles. Instead, this might be interpreted as evidence for a reciprocal intrachromosomal exchange process between the duplicated fragments. However, more detailed analyses using additional human populations as well as additional Y chromosome markers revealed that this phenomenon is highly population-specific and disappears completely when Y-STR diversity is analysed in association with two Y-SNP haplogroups. We found that the diversity of the two DYS385 loci (and other Y-STRs) is highly depending on the haplogroup background, and that equal proportions of both haplogroups in the German sample explains the nearly identical allele frequency distributions at the two DYS385 loci. Thus, we demonstrate here that allele frequency distributions at duplicate loci that are suggestive of intrachromosomal recombination can be explained solely by population history.

European Journal of Human Genetics (2003) 11, 304–314. doi:10.1038/sj.ejhg.5200960

Keywords: intra-chromosomal recombination; Y chromosome diversity; microsatellites; short tandem repeats; population history; DYS385

Introduction

The human Y chromosome displays an extraordinary amount and variety of different classes of repetitive DNA when compared to other human chromosomes.^{1–4} This is assumed to have arisen as a result of the lack of homologous interchromosomal recombination over evolu-

tionary time.^{5–7} However, intrachromosomal rearrangements during primate evolution, such as tandem duplications and inversions, may in addition have played an important role in shaping the molecular structure of the human Y chromosome.^{4,8–10} Tandem duplications are potential targets for homologous intrachromosomal recombination, and the occurrence of such rearrangements between Y-chromosomal repeats, which can result in sequence deletions, is thought to be involved in phenotypes such as male infertility.⁹ For example, it has been shown recently that the Y-chromosomal azoospermia factor c (AZFc) region consists of a complex of three

*Correspondence: Dr M Kayser, Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher platz 6, D-04103 Leipzig, Germany. Tel: +49 341 9952506; Fax: +49 341 9952555; E-mail: kayser@eva.mpg.de

Received 30 July 2002; revised 6 January 2003; accepted 10 January 2003

palindromes constructed from nearly identical copies, which may have resulted from tandem duplication and inversion.⁹ Intrachromosomal exchange between Y-chromosomal copies derived from the HERV15 class of endogenous retroviruses has also been shown to cause deletions at the AZFa region responsible for male infertility.^{11,12} Although recombination between segmental duplications also occurs in autosomes and can cause genomic instability and disease,^{13,14} the Y chromosome is especially suitable for investigating intrachromosomal rearrangements, because of the absence of interchromosomal processes in the nonrecombining region (NRY).

Investigation of intrachromosomal recombination between tandem duplications on the Y chromosome is usually carried out by large-scale DNA sequence analysis; however, highly polymorphic DNA markers that occur within and are variable between the different intrachromosomal copies should also be useful. A class of DNA markers potentially suitable for this application are microsatellites or short tandem repeat (STR) loci on the Y chromosome, since they display a high level of polymorphism and are widely used to characterize individual male lineages in forensics and molecular anthropology.^{15–17} Another prerequisite for their potential use as markers for studying recombination is that some of them occur in multiple copies on the Y chromosome. Of the less than 30 variable Y-STRs described to date, only four consist of two male-specific alleles, one of which is DYS385.^{15,18} The Y-STR system DYS385 shows one of the highest levels of polymorphism yet observed for Y-STRs, and hence it is often used in forensic^{15,19–23} and population history studies.^{17,24,25} It has been previously assumed that DYS385 exists in two polymorphic copies on the Y chromosome,¹⁵ although the lack of sequence information excluded separate analysis of the assumed two copies.¹⁸ Consequently, no individual locus assignment has been possible, and so the two alleles have been treated as a combined allele class, resulting in a loss of information and potential misinterpretation of true ancestry.

To investigate potential intrachromosomal recombination events on the human Y chromosome and to enhance the utility of DYS385 in forensic and population history studies, we analysed the two loci of DYS385 separately by making use of the accumulated sequence data for the human Y chromosome in publicly available databases. We observed in a German population sample ($n=70$) almost identical allele distributions at the two duplicated DYS385 loci, which could be interpreted as the result of a reciprocal genetic exchange, that is, by an intrachromosomal recombination mechanism. However, further studies of 261 male individuals from five populations, as well as Y chromosome haplotype analysis using SNP and additional STR markers, provide evidence that male population history, rather than a genetic exchange process, is the likely explanation for the observed phenomenon.

Material and methods

Human samples and Y-STR analysis

DNA samples from unrelated males from Germany (Berlin-Brandenburg and Leipzig areas; $n=70$), Poland (Wroclaw area, $n=47$), Spain (Grenada area, $n=49$), Papua New Guinea (highland and coastal areas, $n=48$) and from Cameroon (Yaoundé area, $n=47$) were analysed for the DYS385a and DYS385b loci. For most of the samples, Y-STR data for the loci DYS19 (synonym DYS394), DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393 were already available from our previous studies,^{17,26,27} and for the remaining samples Y-STRs were analysed as described in detail on the website of the Y chromosome Haplotype Reference Database (YHRD) for US Populations (<http://www.ystr.org/usa>).²³ Y-STR nomenclature is according to Kayser *et al.*¹⁵ except that DYS389II is considered without variation at DYS389I.

Separate DYS385a and DYS385b analysis

Sequence information for the two loci DYS385a and DYS385b was obtained from the clones RP11-569J3 (GenBank Accession no. AC022486) and RP11-143C1 (GenBank Accession no. AC007379). We defined the DYS385 repeat of RP11-143C1 as DYS385a and the corresponding repeat of RP11-569J3 as DYS385b. The alignment of 1.0 kb of the flanking part showed a variable sequence starting upstream to position 10930 (RP11-143C1) and downstream to position 127415 (RP11-569J3). We designed two new forward primers so that each locus could be amplified individually in a separate reaction: for DYS385a primer 5'-TGTTGTTGACTGTGGTAGGTA-3' (sequence position 10875–10895 of RP11-143C1), and for DYS385b primer 5'-GGGAATGCAATTTCCCTTTAT3' (reverse complement to sequence 127434–127453 of RP11-569J3). These two primers were designed from the single copy sequence between the two 190 kb copies. Analysis of the two DYS385 loci, DYS385a and DYS385b, was performed in a nested PCR approach: in the first reaction both loci were amplified separately using the two newly designed forward primers DYS385a or DYS385b, respectively, and the DYS385.2B primer reported elsewhere,¹⁸ generating a single product of about 700 bp in length for each Y-STR locus. The PCR reaction contained 1 × GeneAmp PCR buffer II (PE Biosystems), 3.0 mM MgCl₂, 0.2 mM dNTPs (Pharmacia Biotech), 0.2 μM each primer (Metabion), 2.5 U AmpliTaq DNA polymerase (PE Biosystems) and 20–50 ng genomic DNA. The PCR was performed as described elsewhere.¹⁸ In the second PCR, we used the published primers DYS385.1 (5' labelled with TAMRA) and DYS385.2B¹⁸ to obtain a shorter fragment of about 250–280 bp in length, in order to be able to detect the Y-STR polymorphism in the number of 4 bp repeats. PCR conditions were identical with the first PCR (with the exception of the primers). A measure of 2 μl of 1:3–1:5 diluted fluorescent-labelled PCR products of DYS385a and DYS385b were separated on a 5% Long

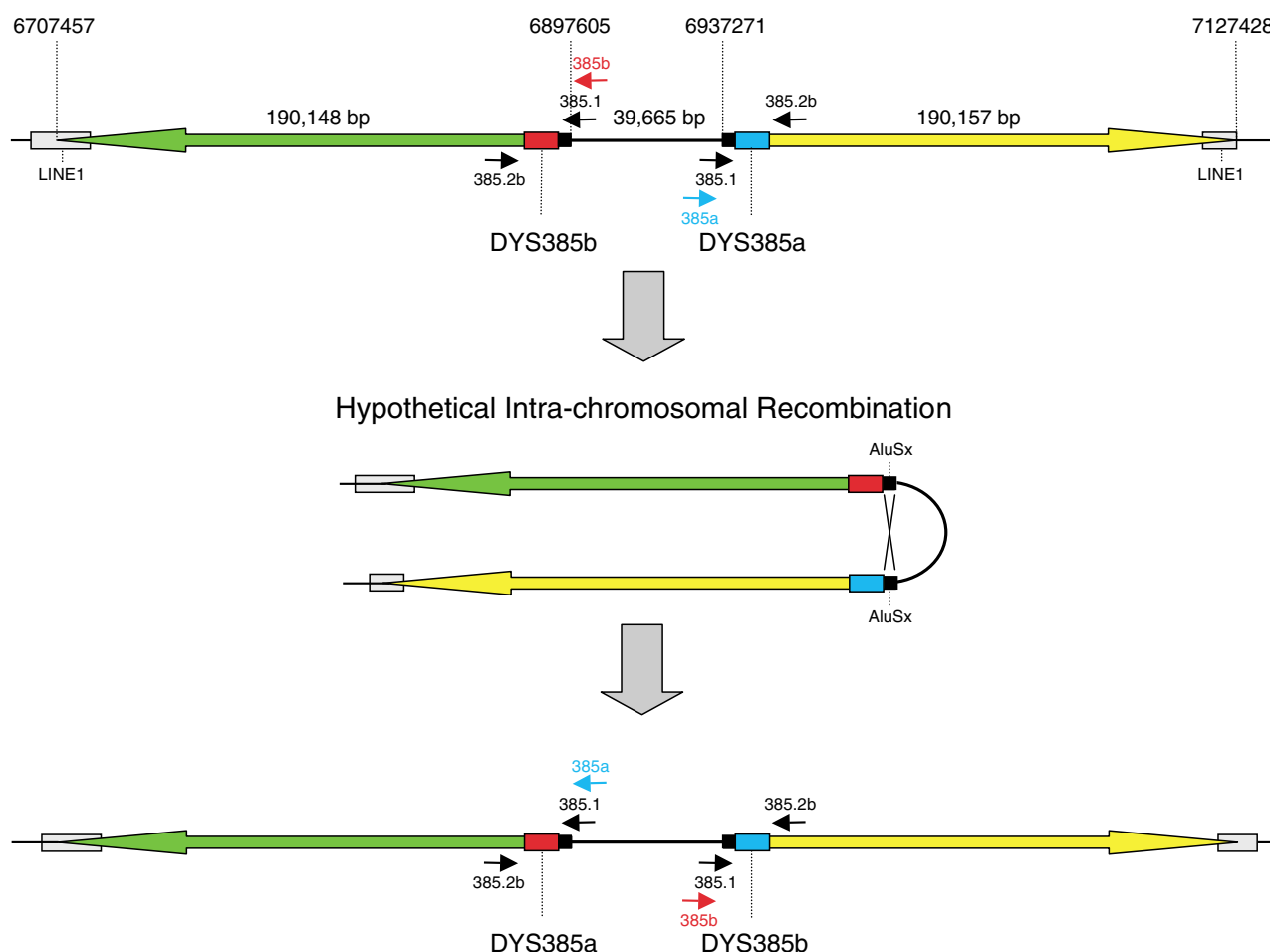


Figure 1 Schematic structure of the inverted repeat containing the Y-chromosomal microsatellite DYS385. The green and yellow arrows represent the duplicated ~ 190 kb repeats, separated by ~ 40 kb of unique sequence, with the sequence orientation indicated. The distal and proximal parts of the repeat are characterized by LINE1 and AluSx sequences. The positions of the distal and proximal boundaries of the repeats are given as sequence positions in the GenBank contig NT_011875 (version NT_011875.7, GI:16172964). The position of the PCR-primers used for specific amplification of the DYS385a and DYS385b microsatellites separately in a nested PCR approach are indicated. The hypothetical intrachromosomal recombination mechanisms resulting in reciprocal genetic exchange is demonstrated.

Ranger™ gel (FMC Bioproducts) using an ABI Prism™ 377 DNA Sequencer (Applied Biosystems, Foster City, CA, USA), and fragment length analysis was performed using the ABI Prism™ Genescan software (Applied Biosystems). DNA sequencing of various length alleles of both loci was performed on ABI Prism™ 377 DNA Sequencer (Applied Biosystems) using Big Dye chemistry (Applied Biosystems, Foster City, CA, USA) according to the recommendations of the manufacturer. For DYS385a, five different length alleles were sequenced (11, 12, 13, 16, 17), and for DYS385b eight alleles were sequenced (11, 12, 14–18, 20). For both loci, allele lengths as obtained from DNA sequencing analysis were 252 bp for allele 10 [(GAAA)₁₀] and 284 bp for allele 18 [(GAAA)₁₈], although alleles always appeared 6 bp shorter in the Genescan fragment length analysis.

Y-chromosomal SNP analysis

Three Y-chromosomal SNPs (SRY-1532, originally called SRY₁₀₈₃₁,^{28,29} 92R7,³⁰ and M9³¹) were analysed in 70 German, 47 Polish and 49 Spanish individuals. Haplogroup 1 (hg1) is characterized by the derived states at SRY-1532 (G), 92R7 (T) and M9 (G); haplogroup 3 (hg3) is characterized by a reversion from the derived G allele to the ancestral A allele at SRY and the derived states at 92R7 and M9.^{32–34} SRY-1532 was genotyped using 0.4 μ M each of the primers 5'-CCTCTGTATCTGACTTTTTCAC-3' and 5'-CCACATAGGTGAACCTTGAAAAT-3', 0.2 mM dNTPs (Pharmacia Biotech), 1 \times GeneAmp PCR buffer II (PE Biosystems), 2.0 mM MgCl₂ (PE Biosystems), and 1 U of AmpliTaq DNA polymerase (PE Biosystems) in a 25 μ l reaction volume. Hot-start PCR (4 min at 95°C) was performed for 35 cycles at

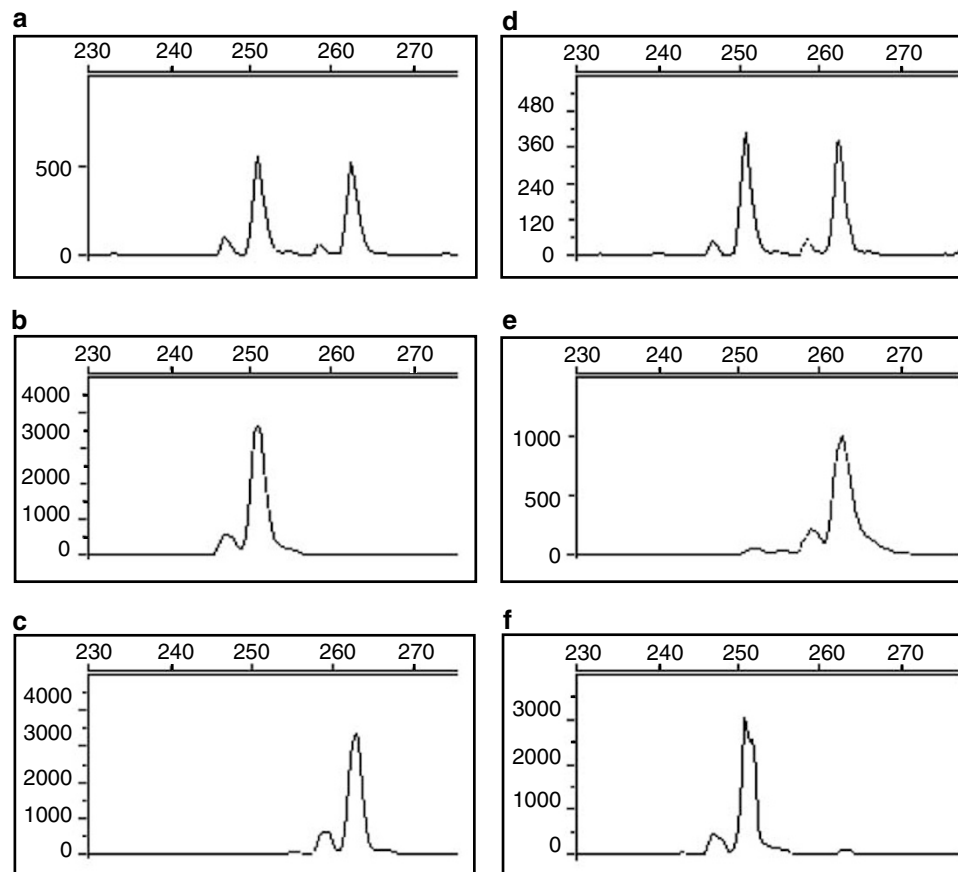


Figure 2 Separate amplification of the two duplicated loci of DYS385. Electropherograms from two individuals with the DYS385a/DYS385b haplotype 11/14 (a–c) and 14/11 (d–f). (a, d): Conventional DYS385 amplification with the primer pair DYS385.1 and DYS385.2B, resulting in two products; (b, e): separate amplification of DYS385a; (c, f): separate amplification of DYS385b. The y-axis represents arbitrary fluorescent units, and the x-axis represents the size of fragments in base pairs.

94°C for 30 s 57°C for 30 s and 72°C for 45 s. SRY-1532 was typed via PCR–RFLP using the restriction endonuclease *Dra*III (New England Biolabs) with the G allele indicated by two fragments of 30 and 26 bp and the A allele by one (undigested) fragment of 56 bp. 92R7 was analysed according to Hurles *et al*³⁵ and M9 as described elsewhere.³⁶

Statistical and computational analyses

Sequence blast search was performed using BLASTN software (version 2.2.1) against the NCBI *Homo sapiens* genomic contig sequences database (<http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs>). LINE-1 and Alu elements were identified using the RepeatMasker software (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>). Correlation analysis between allele frequency distributions was performed using the software package Statistica (Statsoft). A χ^2 -test for independence between allele frequency distributions and respective p-values from Monte Carlo simulations was performed using the program StatXact. The diversity of Y-STR haplotypes, according to Nei,³⁷ was calculated using the software package Arlequin.³⁸

Results and discussion

The DNA sequence of DYS385 was searched against the NCBI database of *Homo sapiens* genomic contig sequences. By large-scale alignment of the identified flanking sequence of DYS385, we detected an ~190 kb fragment, which exists in two copies separated by ~40 kb of unique sequence, mapping to Yq11.223. Sequence identity of the two copies, excluding repetitive sequence variation, was estimated to be 99.97%, which is identical to that estimated for the different copies of a 3 Mb tandem duplication in the AZFc region of the human Y chromosome.⁹ The two ~190 kb copies are located in opposite orientations and thus contain the DYS385 microsatellite in their most proximal parts (Figure 1). The duplication event is likely to have occurred prior to the emergence of anatomically modern humans (100 000–200 000 years ago), since two male-specific alleles at DYS385 have been observed in all human populations studied to date.^{15,17,18,20,22,24,25,39–44}

We used the sequence information to design PCR assays specific for each individual DYS385 locus, designated

DYS385a and DYS385b (Figure 1). Separate and locus-specific amplification of DYS385a and DYS385b was assured by using part of the ~ 40 kb single-copy sequence between the two ~ 190 kb fragments for specific DYS385a and DYS385b primer design (Figure 1). In all separate analyses of the DYS385a and DYS385b loci, a single male-specific PCR-product was observed, compared to the typically two-band pattern resulting from the simultaneous amplification of both loci following the conventional DYS385 PCR protocol (Figure 2). DNA samples, which produce a one-band pattern in the conventional PCR, yielded as expected two fragments of identical size from the analysis of the separate loci, confirming that single-band patterns for DYS385 in the conventional approach do indeed reflect two alleles of identical length at the two loci, as assumed previously,^{15,18} and not the existence of only one locus. DNA sequence analysis of several alleles of varying lengths at both DYS385 loci revealed no sequence differences either within or between the two loci, except for the expected variation in the number of (GAAA)_n repeats.

The separate analysis of DYS385a and DYS385b in 70 unrelated German males revealed an unexpected and striking similarity of the allele frequency distributions at both Y-STR loci (Figure 3a). The two allele frequency distributions are significantly correlated ($r=0.853$; $P<0.001$) and are not significantly different from one another by a χ^2 -test ($\chi^2=15.45$; $df=10$; $P>0.05$) (Table 1). Such a strong similarity of allele frequency distributions of two duplicated STR loci cannot be explained by microsatellite DNA evolution according to the generally assumed polymerase/strand slippage model,^{45,46} for the following reason. The duplication event is old, perhaps as old as 100 000–200 000 years, because all human populations carry two copies of the DYS385 microsatellite. After the duplication event, both copies of the DYS385 microsatellite started to mutate independently from each other as they are separated by ~ 40 kb of interspersed unique sequence. Given a mutation rate of about 2×10^{-3} per generation as experimentally observed for DYS385,²⁶ it is very unlikely that the microsatellite evolution would by chance result in

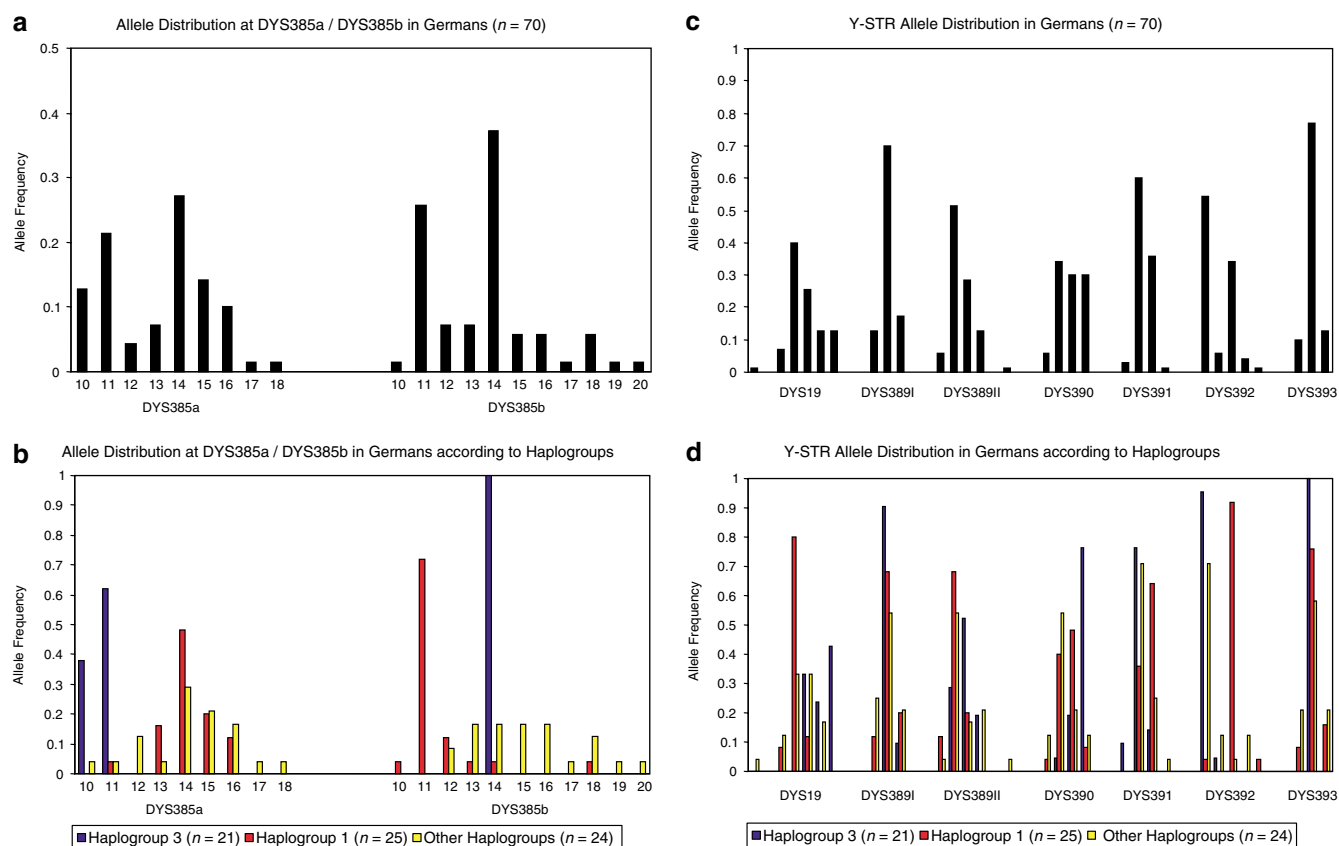


Figure 3 Allele frequency distributions at DYS385a and DYS385b, and at seven additional Y-STRs, for a different German (a–d), Polish (e–h) and Spanish (i–l) population sample. Allele frequency distributions are presented both for all pooled individuals and separately according to Y-SNP haplogroups.

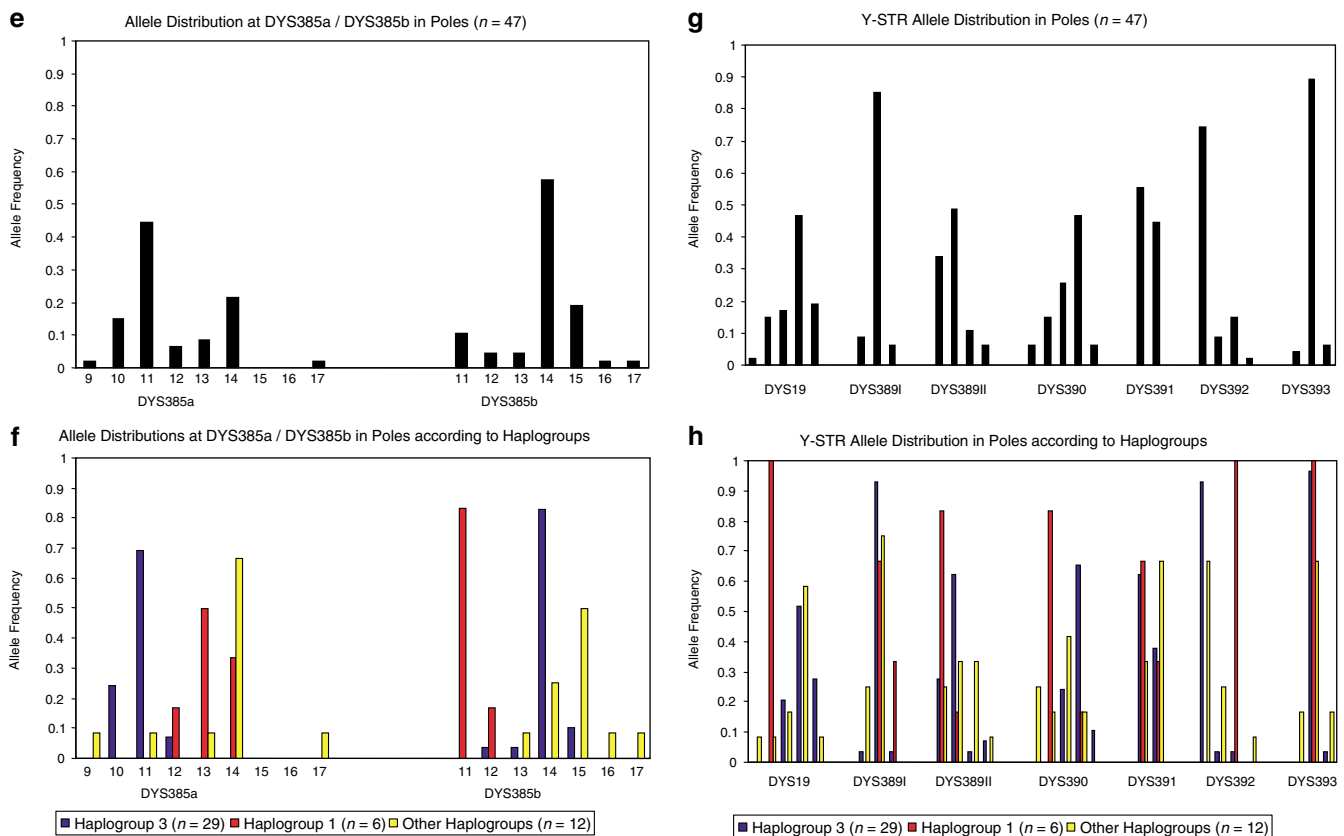


Figure 3 Continued

almost identical, bimodal allele frequency distributions at both loci.

One way to explain such a nearly identical allele frequency distribution at two independently mutating Y-chromosomal microsatellite loci would be intrachromosomal genetic recombination. Such events would result in reciprocal exchange of sequence information between the two duplicated 190 kb fragments and thus, if frequent enough, would make the allele frequency distribution at the two copies of the DYS385 microsatellite almost identical (Figure 1). Since we detect the two DYS385 copies by specific PCR primers located outside the 190 kb fragment, the putative intrachromosomal recombination events must occur in the most proximal part of the 190 kb fragments to result in a complete reciprocal exchange including the DYS385 microsatellite (Figure 1).

It should be emphasized that nonreciprocal events such as gene conversion would result in different expectations than reciprocal genetic exchange. Gene conversion or other nonreciprocal events would create identical copies at the two loci, and thus an excess of homozygotes and a significant correlation of DYS385a and DYS385b at the individual level. Neither of these was observed in our data.

If intrachromosomal recombination is indeed the underlying cause for the similar DYS385a and DYS385b allele frequency distributions in the German sample, then we would expect other populations to also exhibit similar allele frequency distributions. We thus genotyped DYS385a and DYS385b in additional human populations from Poland, Spain, Papua New Guinea and Cameroon. However, all these other groups showed considerably different allele distributions between the two DYS385 loci, with nonsignificant correlation coefficients and with significant χ^2 values (Table 1). These results indicate that instead of a general mechanism such as intrachromosomal recombination, a different phenomenon that exhibits population-specificity is apparently responsible for the strong similarity between the two allele frequency distributions in the German sample.

To investigate how the paternal history of Germans might result in similar allele frequency distributions at DYS385a and DYS385b, we genotyped the German, Polish and Spanish samples for three Y-SNPs, which define two Y chromosome haplogroups (hg1 and hg3) known from previous studies to be informative in European populations.^{32–35,47} The highly significant correlation in the German sample between the allele frequency distributions

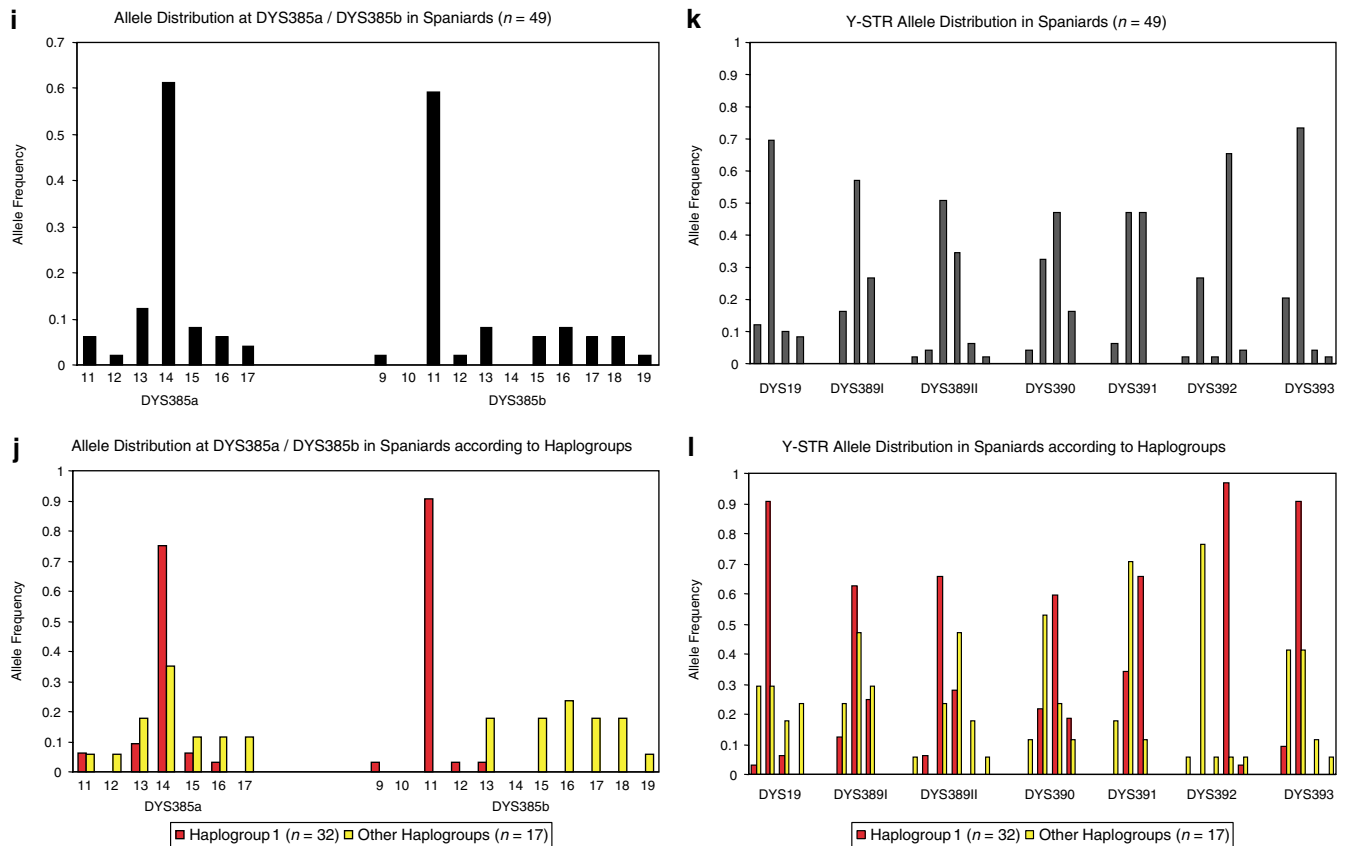


Figure 3 Continued

Table 1 Allele frequencies (%) at DYS385a and DYS385b (analysed separately) in different human populations

	Germans (n=70)		Poles (n=47)		Spaniards (n=49)		PNG (n=48)		Cameroon (n=47)	
Allele	DYS385a	DYS385b	DYS385a	DYS385b	DYS385a	DYS385b	DYS385a	DYS385b	DYS385a	DYS385b
9	9 (12.9)	1 (1.4)	1 (2.1)	0	0	1 (2.0)	1 (2.1)	1 (2.1)	0	0
10	15 (21.4)	18 (25.7)	7 (14.9)	0	0	0	1 (2.1)	0	0	0
11	3 (4.3)	5 (7.1)	21 (44.7)	5 (10.6)	3 (5.9)	29 (56.9)	4 (8.3)	1 (2.1)	0	0
12	5 (7.1)	5 (7.1)	3 (6.4)	2 (4.3)	1 (2.0)	1 (2.0)	13 (27.1)	10 (20.8)	0	0
13	19 (27.1)	26 (37.1)	4 (8.5)	2 (4.3)	6 (12.2)	4 (7.8)	16 (33.3)	2 (4.2)	0	0
14	10 (14.3)	4 (5.7)	10 (21.3)	27 (57.4)	30 (61.2)	0	6 (12.5)	4 (8.3)	0	0
15	7 (10.0)	4 (5.7)	0	9 (19.1)	4 (7.8)	3 (5.9)	6 (12.5)	9 (18.8)	41 (85.4)	1 (2.1)
16	1 (1.4)	1 (1.4)	0	1 (2.1)	3 (5.9)	4 (7.8)	1 (2.1)	6 (12.5)	2 (4.2)	36 (75.0)
17	1 (1.4)	4 (5.7)	1 (2.1)	1 (2.1)	2 (3.9)	3 (5.9)	0	2 (4.2)	3 (6.3)	4 (8.3)
18	0	1 (1.4)	0	0	0	3 (5.9)	0	5 (10.4)	1 (2.1)	5 (10.4)
19	0	1 (1.4)	0	0	0	1 (2.0)	0	2 (4.2)	0	1 (2.1)
20	0	0	0	0	0	0	0	5 (10.4)	0	0
21	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	1 (2.1)	0	0
r^a	0.853 ($P < 0.001$)		0.306 ($P > 0.05$)		-0.106 ($P > 0.05$)		0.399 ($P > 0.05$)		-0.294 ($P > 0.05$)	
χ^2_{b}	15.45 ($P > 0.05$)		36.52 ($P < 0.001$)		57.01 ($P < 0.001$)		33.65 ($P < 0.001$)		72.33 ($P < 0.001$)	

^aCorrelation coefficient and P -value between the allele frequency distributions for DYS385a and DYS385b, bold values are statistically significant. ^b χ^2 -value and P -value between the allele frequency distributions for DYS385a and DYS385b, bold values are statistically significant; allele nomenclature as described elsewhere.¹⁸

Table 2 Correlation coefficients (*r*) and their *P*-values^a between the allele frequency distributions for DYS385a and DYS385b

	Germans		Poles		Spaniards	
	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>	<i>r</i>	<i>P</i>
All individuals	0.853	0.00085	0.306	0.423	−0.106	0.757
Haplogroup 1	−0.193	0.618	−0.868	0.132	−0.116	0.784
Haplogroup 3	−0.390	0.516	−0.370	0.471	—	—
Pooled haplogroups 1 and 3	0.819	0.007	−0.370	0.471	−0.116	0.785
Other haplogroups	0.679	0.022	0.245	0.525	−0.131	0.736

^aStatistically significant *r*-values indicated in bold.

for DYS385a and DYS385b disappeared completely when the variation was partitioned by the two Y-SNP haplogroups (Table 2, Figure 3a, b). In those individuals, which belonged to neither hg1 nor hg3 (and thus are potentially a pool of different Y-SNP haplogroups), the correlation coefficient was still statistically significant ($r=0.679$, $P=0.022$), albeit lower than in the entire sample (Table 2, Figure 3b). This indicates that differences in the allele frequency distributions according to the Y-SNP background, primarily (but not exclusively) based on hg1 and hg3, are responsible for the strong similarity in allele frequency distribution between DYS385a and DYS385b in the German sample.

More specifically, the combined Y-SNP/Y-STR analysis revealed that there are two alleles at DYS385a and DYS385b that are strongly correlated with the SNP background in the German sample, but in opposite directions. All DYS385a/DYS385b 11/14 haplotypes belonged to hg3, whereas all 14/11 haplotypes belonged to hg1. This relation is also evident from the individual allele frequency distributions for DYS385a and DYS385b, separated according to haplogroup (Figure 3b). Although there is no significant correlation between the DYS385a and DYS385b allele frequency distributions in the Polish and Spanish samples (Table 2, Figure 3e, f, i, j), significant differences are observed in the distribution of DYS385a and DYS385b alleles according to the two Y-SNP haplogroups in all European groups analysed. In particular, alleles DYS385a-11 and DYS385b-14 are associated with hg3, and DYS385a-14 and DYS385b-11 with hg1.

Striking differences between the allele frequency distributions for the different Y-SNP haplogroups were also observed at nearly all other Y-STR loci analysed in all three populations (Figure 3c/d, g/h, k/l). In particular, at DYS392 nearly all hg3 individuals carry allele 11, whereas nearly all hg1 individuals carry allele 13. Since hg3 and hg1 are assumed to be monophyletic, consequently there was a single ancestral Y chromosome, with a specific Y-STR haplotype, for each haplogroup. Our data suggest that for hg3 this ancestral Y-STR haplotype was DYS19-389I-389II-390-391-392-393-385a-385b = ?-10-17-25-10-11-13-11-14, and for hg1 it was 14-10-16-24(23)-11(10)-13-13-14-11.

Differences in the amount of variation at each Y-STR locus, within each haplogroup, can partly be explained by variation in the locus-specific mutation rates. DYS392 and DYS393 showed very low levels of variation within the haplogroups (Figure 3) and also the lowest mutation rate among all Y-STR loci used here, whereas mutation rates for the other Y-STR loci were much higher.²⁶ Our observations clearly illustrate that Y-STR haplotype variation is strongly dependent on the relevant Y-SNP background, confirming previous findings at single Y-STR loci.^{33,48,49}

Significant associations of particular Y-STR haplotypes with particular haplogroups is not unexpected, but how does this explain nearly identical allele frequency distributions for DYS385a and DYS385b in the German sample, but no other sample? The answer apparently lies in different frequency distributions of the Y-SNP haplogroups, together with associated Y-STR alleles, resulting from different paternal histories of these populations. The frequencies of hg1 and hg3 differ significantly between the German, Polish and Spanish samples ($\chi^2=48.77$, $df=4$, $P<0.0001$). In the German sample, both haplogroups occur in similar frequency, whereas in the Polish sample the proportion of hg3:hg1 is about 5:1, and in the Spanish sample only hg1 exists with hg3 being totally absent (Table 3). It has been shown recently that hg3 and hg1 display a clinal distribution with increasing frequency for hg1 from the southwest of Europe to the northeast, and for hg3 from the central-eastern part of Europe to the southeast and southwest.³⁴ Hg3 is assumed to have an eastern European origin about 7500⁵⁰ or 2550–3800 years BP.^{33,34} The higher frequency of hg3 in our Polish sample as well as a higher Y-STR haplotype diversity associated with hg3 in the Polish sample than in the German sample (Table 3) and the absence of hg3 in the Spanish sample, are in agreement with this presumed eastern origin of hg3.³⁴ Hg1 is hypothesized to reflect the predominant early European Y chromosome lineage, which was widely distributed before being partially replaced by a major demographic expansion of agricultural migrants from the Near East into Europe carrying Y chromosome haplogroup hg9.³⁴ Our observation of a higher frequency of hg1 in the German sample compared to the Polish sample, and especially with

Table 3 Frequency of Y-SNP haplogroups and associated Y-STR haplotype^a diversity in the German, Polish and Spanish samples

	Y-SNP haplogroup frequency (%)	Number of Y-STR haplotypes	Y-STR haplotype diversity \pm SE
<i>Germans (n=70)</i>			
Haplogroup 1	25 (35.7)	24	0.997 ± 0.013
Haplogroup 3	21 (30.0)	17	0.976 ± 0.023
Other haplogroups	24 (34.3)	24	1.000 ± 0.012
<i>Poles (n=47)</i>			
Haplogroup 1	6 (12.8)	6	1.000 ± 0.096
Haplogroup 3	29 (61.7)	24	0.985 ± 0.014
Other haplogroups	12 (25.5)	11	0.985 ± 0.040
<i>Spaniards (n=49)</i>			
Haplogroup 1	32 (65.3)	28	0.980 ± 0.019
Haplogroup 3	0	0	0
Other haplogroups	17 (34.7)	17	1.000 ± 0.020
<i>All samples (n=166)</i>			
Haplogroup 1	63 (37.9)	54	0.992 ± 0.006
Haplogroup 3	50 (30.1)	36	0.980 ± 0.012
Other haplogroups	53 (31.9)	52	0.999 ± 0.005

^aY-STR haplotypes consist of nine loci: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b.

the highest frequency observed in the Spanish sample (Table 3), fits this hypothesis. The SNP mutation defining hg1 is assumed to be rather old, for example, ~23 000 years BP⁵⁰ and the high Y-STR-haplotype diversity observed here for all three populations presumably reflects the old age of this haplogroup (Table 3).

Thus, the chance 1:1 ratio of hg1 and hg3, together with their respectively associated DYS385 alleles, appears to be responsible for the nearly identical allele frequency distributions of DYS385a and DYS385b in the German sample. Proceeding from Germany to the east, to Poland, shifts the proportion of hg3:hg1 towards hg3 owing to its eastern European origin, which results in the disappearance of the similar DYS385a and DYS385b allele frequency distributions. Similarly, proceeding from Germany to the west, to Spain, shifts the 1:1 ratio of hg3 and hg1 completely towards hg1, thereby also resulting in different DYS385a and DYS385b allele frequency distributions. That we see in the German sample a lower but still statistically significant correlation when hg1 and hg3 are excluded indicates that other haplogroups and associated Y-STR haplotypes also contribute to this effect.

Consequently, the amount of incorrect allele–locus assignment for DYS385 when analysed simultaneously following the conventional DYS385 PCR protocol strongly depends on the population investigated and their hg1:hg3 proportion. When analysed simultaneously, the convention for statistical analyses of DYS385 data is that the shorter allele is artificially assigned to DYS385a and the longer allele to DYS385b. Thus, for hg3 Y chromosomes, with high frequency of DYS385a-11 and DYS385b-14 (and their derivatives), this approach would reflect the true

allele–locus assignment, and in general the higher the frequency of hg3 chromosomes in a population, the lower the misclassification rate. In contrast, for hg1 Y chromosomes, with high frequencies of DYS385a-14 and DYS385b-11 (and their derivatives), this approach would not reflect the true allele–locus assignment, and the higher the frequency of hg1 chromosomes in a population, the higher the misclassification rate. In our data, the amount of incorrect DYS385 assignment was 40% in the German sample, decreasing to 12.8% in the Polish sample, but increasing to 65.3% in the Spanish sample, which exactly mirrors the relative proportion of hg1 chromosomes. This might be crucial for forensic applications of DYS385 using the traditional simultaneous amplification approach, especially if the number of additional Y-STR loci analysed is low.

Conclusion

We unexpectedly observed almost identical allele frequency distribution at two copies of the duplicated Y-STR locus DYS385 in a German sample. Although we initially suspected that intrachromosomal recombination could be responsible, further investigation showed that this phenomenon was restricted to the German sample. Investigation of the Y-SNP background revealed that the identical DYS385 allele frequency distributions in the German sample disappeared when the variation was partitioned by Y-SNP haplogroups. We showed that Y-STR variation depends strongly on the Y-SNP background; in particular two Y-SNP haplogroups (hg1 and hg3) are associated with complementary genotypes at the two DYS385 loci (14/11 and 11/14, respectively). These two haplogroups occur in

roughly equal frequency in the German sample, but in different proportions in other samples, explaining the initial finding of identical allele frequency distributions at the two DYS385 loci in the German sample but not other samples. Thus, our data demonstrate that nearly identical allele frequency distributions at two highly polymorphic STR loci embedded in a large duplicated DNA-fragment on the Y chromosome, which may be interpreted as result of reciprocal genetic exchange (i.e. intrachromosomal recombination), can rather be an effect caused solely by population history. This should be taken into consideration when dealing with putative evidence for intrachromosomal recombination, especially on the Y chromosome. Furthermore, our results show that individuals with similar phenotypes at duplicate Y-STR loci can have different genotypes, which has important implications for forensic but also evolutionary applications involving such loci.

Acknowledgements

We thank Christiane Rammelt for technical assistance. Lutz Roewer, Tadeusz Dobosz, Miguel Lorente and Pete Zimmerman are gratefully acknowledged for DNA samples. We thank Wolfgang Enard, Michael Hofreiter, Chris Tyler-Smith, Linda Vigilant and Gunter Weiss for useful comments on the manuscript. RK and AE were supported by the German National Merit Foundation (Studienstiftung des Deutschen Volkes). This study was funded by the Max Planck Society.

References

- 1 Foote S, Vollrath D, Hilton A, Page DC: The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* 1992; **258**: 60–66.
- 2 Vollrath D, Foote S, Hilton A *et al*: The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* 1992; **258**: 52–59.
- 3 Erlandsson R, Wilson J, Pääbo S: Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. *Mol Biol Evol* 2000; **17**: 804–812.
- 4 Tilford CA, Kuroda-Kawaguchi T, Skaletsky H *et al*: A physical map of the human Y chromosome. *Nature* 2001; **409**: 943–945.
- 5 Charlesworth B, Langley CH, Stephan W: The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* 1986; **112**: 947–962.
- 6 Charlesworth B: The evolution of sex chromosomes. *Science* 1991; **125**: 1030–1033.
- 7 Charlesworth B, Sniegowski P, Stephan W: The evolutionary dynamics of repetitive DNA in eucaryotes. *Nature* 1994; **371**: 215–220.
- 8 Lahn BT, Page DC: Four evolutionary strata on the human X chromosome. *Science* 1999; **286**: 964–967.
- 9 Kuroda-Kawaguchi T, Skaletsky H, Brown LG *et al*: The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* 2001; **29**: 279–286.
- 10 Samonte RV, Eichler EE: Segmental duplications and the evolution of primate genome. *Nat Rev Genet* 2001; **3**: 65–72.
- 11 Sun C, Skaletsky H, Rozen S *et al*: Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum Mol Genet* 2000; **9**: 2291–2296.
- 12 Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME: Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J Med Genet* 2000; **37**: 752–758.
- 13 Osborne LR, Li M, Pober B *et al*: A 1.5 million-base pair inversion polymorphism in families with Williams–Beuren syndrome. *Nat Genet* 2001; **29**: 321–325.
- 14 Emanuel BS, Shaikh TH: Segmental duplications: an ‘expanding’ role in genomic instability and disease. *Nat Rev Genet* 2001; **2**: 791–800.
- 15 Kayser M, Caglia A, Corach D *et al*: Evaluation of Y chromosomal STRs: a multicenter study. *Int J Legal Med* 1997; **110**: 125–133, 141–149.
- 16 Kayser M, Krawczak M, Excoffier L *et al*: Extensive analysis of chromosome Y microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* 2001; **68**: 990–1018.
- 17 Kayser M, Brauer S, Schädlich H *et al*: Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. *Genome Res* 2003; in press.
- 18 Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C: Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci Int* 1998; **97**: 61–70.
- 19 Corach D, Sala A, Penacino G *et al*: Additional approaches to DNA typing of skeletal remains: the search for missing persons killed during the last dictatorship in Argentina. *Electrophoresis* 1997; **18**: 1608–1612.
- 20 Caglia A, Dobosz M, Boschi I, d’Aloja E, Pascali VL: Increased forensic efficiency of a STR-based Y-specific haplotype by addition of the highly polymorphic DYS385 locus. *Int J Legal Med* 1998; **111**: 142–146.
- 21 Schneider PM, d’Aloja E, Dupuy BM *et al*: Results of collaborative study regarding the standardization of the Y-linked STR system DYS385 by the European DNA Profiling (EDNAP) group. *Forensic Sci Int* 1999; **102**: 159–165.
- 22 Roewer L, Krawczak M, Willuweit S *et al*: Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int* 2001; **118**: 106–113.
- 23 Kayser M, Brauer S, Willuweit S *et al*: Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations. *J Forensic Sci* 2002; **47**: 513–519.
- 24 Seielstad M, Bekele E, Ibrahim M, Toure A, Traore M: A view of modern human origins from Y chromosome microsatellite variation. *Genome Res* 1999; **9**: 558–567.
- 25 Helgason A, Sigurðardóttir S, Nicholson J *et al*: Estimating Scandinavian and Gaelic Ancestry in the male settlers of Iceland. *Am J Hum Genet* 2000; **67**: 697–717.
- 26 Kayser M, Roewer L, Hedman M *et al*: Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 2000; **66**: 1580–1588.
- 27 Kayser M, Brauer S, Weiss G, Schiefelhövel W, Underhill P, Stoneking M: Independent histories of human Y chromosomes from Melanesia and Australia. *Am J Hum Genet* 2001; **68**: 173–190.
- 28 Whitfield LS, Sulston JE, Goodfellow PN: Sequence variation of the human Y chromosome. *Nature* 1995; **378**: 379–380.
- 29 Kwok C, Tyler-Smith C, Medonca BB *et al*: Mutation analysis of 2 kb 5’ to SRY in XY females and XX intersex subjects. *J Med Genet* 1996; **33**: 465–468.
- 30 Mathias N, Bayes M, Tyler-Smith C: Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 1994; **3**: 115–123.
- 31 Underhill PA, Jin L, Lin AA *et al*: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; **7**: 996–1005.
- 32 Tyler-Smith C: Y-chromosomal DNA markers; in Papiha SS, Deka R, Chakraborty R (eds): *Genomic diversity: applications in human population genetics*. New York: Kluwer Academic/Plenum Publishers, 1999, pp 65–73.
- 33 Zerjal T, Pandya A, Santos FR *et al*: The use of Y-chromosomal DNA variation to investigate population history; in Papiha SS, Deka R, Chakraborty R (eds): *Genomic diversity: applications in*

- human population genetics*. New York: Kluwer Academic/Plenum Publishers, 1999, pp 91–101.
- 34 Rosser ZH, Zerjal T, Hurles ME *et al*: Y-chromosomal diversity within Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 2000; **67**: 1526–1543.
- 35 Hurles ME, Veitia R, Arroya E *et al*: Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* 1999; **65**: 1437–1448.
- 36 Kayser M, Brauer S, Weiss G *et al*: Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 2000; **10**: 1237–1246.
- 37 Nei M. *Molecular evolutionary genetics*. New York: Columbia University Press, 1987.
- 38 Schneider S, Roessli D, Excoffier L: *Arlequin ver 2.000: a software for population genetics data analysis*. University of Geneva, Switzerland: Genetics and Biometry Laboratory, 2000.
- 39 Füredi S, Woller J, Pádár Z, Angyal M: Y-STR haplotyping in two Hungarian populations. *Int J Legal Med* 1999; **113**: 38–42.
- 40 Miranda JJ, Benecke M, Hidding M, Schmitt C: Y-chromosomal short tandem repeat haplotypes at the loci DYS393, DYS19, DYS392, and DYS385-I/II, DYS390, DYS389-I/II, and DYS391 in a Filipino population sample. *J Forensic Sci* 2001; **46**: 1250–1253.
- 41 Wu FC, Pu CE: Multiplex DNA typing of short tandem repeat loci on Y chromosome of Chinese population in Taiwan. *Forensic Sci Int* 2001; **120**: 213–222.
- 42 Nonaka I, Minaguchi K: Allele frequencies and haplotypes of ten Y-specific STRs in the Japanese population. *J Forensic Sci* 2001; **46**: 179–182.
- 43 Gamero JJ, Romero JL, Gonzalez JL *et al*: Population-genetic study of the DYS385 haplotypes in two Spanish populations and the African immigrant population in Spain. *J Forensic Sci* 2001; **46**: 193.
- 44 Rangel-Villalobos H, Jaloma-Cruz AR, Sandoval-Ramirez L, Velarde-Felix JS, Gallegos-Arreola MP, Figuera LE: Y-chromosome haplotypes for six short tandem repeats (STRs) in a Mexican population. *Arch Med Res* 2001; **32**: 232–237.
- 45 Levidson G, Gutman GA: Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 1987; **4**: 203–221.
- 46 Schlötterer C, Tautz D: Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 1992; **20**: 211–215.
- 47 Santos FR, Tyler-Smith C: Reading the human Y chromosome: the emerging DNA markers and the human genetic history. *Braz J Genet* 1996; **19**: 665–670.
- 48 de Knijff P: Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am J Hum Genet* 2000; **67**: 1055–1061.
- 49 Nebel A, Filon D, Hohoff C, Faerman M, Brinkmann B, Oppenheim A: Haplogroup-specific deviation from the stepwise mutation model at the microsatellite loci DYS388 and DYS392. *Eur J Hum Genet* 2001; **9**: 22–26.
- 50 Karafet TM, Zegura SL, Posukh O *et al*: Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* 1999; **64**: 817–831.