

Variation of 52 new Y-STR loci in the Y Chromosome Consortium worldwide panel of 76 diverse individuals

Si-Keun Lim · Yali Xue · Emma J. Parkin ·
Chris Tyler-Smith

Received: 28 March 2006 / Accepted: 21 August 2006 / Published online: 21 October 2006
© Springer-Verlag 2006

Abstract We have established 16 small multiplex reactions of two–four loci to amplify 52 recently described single-copy simple Y-STRs and typed these loci in a worldwide panel of 74 diverse men and two women. Two Y-STRs were found to be commonly multicopy in this sample set and were excluded from the study. Of the remaining 50, four (DYS481, DYS570, DYS576 and DYS643) showed higher diversities than the commonly used loci and can potentially provide increased haplotype discrimination in both forensic and anthropological work. Ten loci showed occasional missing alleles, duplicated peaks or intermediate-sized alleles.

Keywords Y chromosome · Short tandem repeat (STR) · DYS481 · DYS570 · DYS576 · DYS643 · Intermediate allele

Electronic supplementary material Supplementary material is available for this article at <http://dx.doi.org/10.1007/s00414-006-0124-8> and is accessible for authorized users.

S.-K. Lim · Y. Xue · C. Tyler-Smith (✉)
The Wellcome Trust Sanger Institute,
Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
e-mail: cts@sanger.ac.uk

E. J. Parkin
Department of Genetics, University of Leicester,
Leicester, UK

Present address:

S.-K. Lim
National Institute of Scientific Investigation,
Seoul, South Korea

Introduction

Y-STRs have key roles in the fields of forensic genetics, anthropological genetics and genealogy because of their ability to discriminate between male lineages and provide information about the relationships between them [1, 2]. The Y chromosome haplotype reference database [3] provides a widely used compilation of haplotype information constructed from a “minimal haplotype” of nine loci or a “minHt+SWGDAM core set” of 11 loci (<http://www.yhrd.org/index.html>). Some applications, however, require more Y-STRs. For example, a study of ~1,000 men from east Asia found that almost 3% (27/1,003) shared the same 16-STR haplotype [4] and thus would not be distinguished by standard analyses. Most of the STRs on the Y chromosome have now been identified [5], and a set of 52 was highlighted that seemed particularly useful because their unit size was ≥ 3 , they were single-copy, had a simple structure and showed variation in a set of eight diverse men. These additional loci proved to be useful in the east Asian study where 46 of them allowed a male lineage characteristic of the Qing Dynasty to be defined [4], but they clearly varied considerably in their diversity [4, 5] and may vary in other properties that affect their usefulness as well. In addition, it may often be impractical or impossible to type such a large number of markers. Further studies of these loci are therefore needed to identify the most useful subset. US population data for 16 of them have been presented [6], but data from other loci and populations are lacking. We have therefore established multiplex typing procedures for all of them and examined their variation in the Y Chromosome Consortium (YCC) worldwide panel of men [7].

Materials and methods

The YCC panel consists of 74 male and two female DNAs; the men may be broken down into 26 from Africa, 26 from Asia and the Americas and 22 from Europe or the Middle East. In addition, the haplogroup R individual previously typed with all of the new markers [5] was included in this study to facilitate consistent allele calling. DNA was amplified before use with the GenomiPhi whole genome amplification kit (Amersham Biosciences, Amersham, UK) according to the manufacturer's recommendations.

A total of 52 polymorphic simple single-copy Y-STRs [5] were included in the present study. The published primers had been designed to operate under a common set of conditions and were therefore used in this study, except that a G was added to the 5' end of the unlabelled primer if it was not already present to facilitate non-templated addition of an A to the labelled product strand [8]. Loci were tested *in silico* for potential interactions between primers using the AutoDimer software [9], and suitable sets were assembled into small multiplexes for experimental assessment resulting in 16 multiplexes each consisting of 2–4 loci (Table S1).

Polymerase chain reactions (PCRs) were set up in 20 µl volumes containing 1× PCR buffer (Invitrogen, Paisley, UK), 1.75 mM MgCl₂, 200 µM deoxynucleotide triphosphates (dNTPs; Amersham Biosciences), 1.0 unit of Platinum *Taq* DNA polymerase (5 U/µl, Invitrogen) with 10 pg–2 ng whole-genome-amplified DNA and primer pairs at the concentrations shown in Table S1. Thermal cycling was carried out in an MJ Research (Genetic Research Instrumentation, Braintree, UK) DNA Engine Tetrad™ 2 starting with denaturation at 95°C for 15 min, followed by 20 cycles of touchdown PCR: 94°C for 30 s, 70°C for 45 s, 72°C for 1 min, with a 1°C decrease in annealing temperature every cycle and then 15 cycles of

standard PCR (94°C for 30 s, 50°C for 45 s, 72°C for 1 min) and finishing with extension at 60°C for 45 min and storage at 4°C.

Products were analysed by mixing 1 µl of PCR product with 15 µl Hi-Di formamide and 0.2 µl size marker (CXR 60–400 bases, Promega UK, Southampton, UK) and running on 36 cm×50 µm capillaries containing POP-4 polymer (Applied Biosystems) on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Warrington, UK). Electrophoresis was carried out at 3 kV for 3 s followed by 15 kV for 45 min with a run temperature of 60°C. Allele sizes were measured using GeneMapper v3.0 (Applied Biosystems). Most loci were sequenced because of the lack of previous sequence data, to confirm previous results or to investigate the structure of intermediate-sized sizes. Such alleles were amplified using unlabelled primers and sequenced by the Wellcome Trust Sanger Institute small-scale sequencing facility using standard methods.

Results

The 52 Y-STRs were examined in the 76 YCC samples and haplogroup R control individual, but the analyses presented in this paper (Tables S2, Tables S3) are based only on the YCC data to facilitate comparisons with other YCC results [10]. As expected, no specific products were obtained from the two female YCC samples in the size range examined, and single peaks were seen in all males for 40 of the STRs. The other 12 loci showed more complex patterns (Table 1). Products from four loci were missing in one (DYS525, DYS589, DYS636) or two (DYS556) individuals. These findings were reproducible and occurred in multiplex reactions that successfully amplified other loci, so that they may represent null alleles, but their structural basis remains

Table 1 Loci showing multiple peaks, missing peaks or intermediate alleles

Locus	Intermediate allele ^a	Comments
DYF386S1		Two peaks in many individuals, excluded from analysis
DYF390S1		Two peaks in many individuals, excluded from analysis
DYS448		Two peaks in two individuals
DYS522	10 (U2Ains)	Intermediate-sized allele. This insertion converts the reference sequence, which can be written ATAG ATG (ATAG) ₁₀ into ATAG ATAG (ATAG) ₁₀ , which therefore has 12 copies of the ATAG repeat, but differs from the regular allele 12
DYS525		No product in one individual, two peaks in one individual
DYS531	11 (D6Tins)	Intermediate-sized allele
DYS549		Two peaks in one individual
DYS556		No product in two individuals
DYS567		Two peaks in two individuals
DYS576		Two peaks in two individuals
DYS589		No product in one individual
DYS636		No product in one individual

U upstream, D downstream, ins insertion

^aNomenclature based on the standard recommendations [1]

to be determined, and they were treated conservatively as missing data in our analyses.

Two peaks were observed in many individuals for DYF390S1 and DYF386S1, and we interpreted these as duplicated loci that happened to have the same sized alleles in the small number of individuals examined before [5]; these two STRs were excluded from subsequent analyses. Five loci also showed two peaks of similar height in one (DYS525, DYS549) or two (DYS488, DYS567, DYS576) individuals, which may reflect rare duplications or somatic mutations in the YCC cell lines. In addition, two loci showed fragment sizes that did not fall into the expected size classes: DYS522 in one individual and DYS531 in 11 individuals corresponding precisely to haplogroup Q [7] and thus representing a variant characteristic of this haplogroup. The structural basis of these variants was determined by sequencing and found to arise from insertion events in the flanking sequences between the STRs and the primers (Table 1). Null alleles, occasional duplications and intermediate alleles have been found in the standard Y-STRs [1], and so we concluded that 50 of the 52 new Y-STRs merited further consideration as loci for wider use.

We next examined the variation of these 50 STRs. The number of alleles ranged from two to 11, the diversity from 0.05 to 0.90 and the variance from 0.04 to 7.89 (Table 2). All of these characteristics were correlated, probably because of their common dependence on the repeat count. To interpret the values obtained, we have compared them with published data on the standard single-copy loci in the YCC panel [10]. Of the new loci, four (DYS481, DYS570, DYS576 and DYS643) showed higher diversity than the most variable standard locus DYS390 (diversity=0.79) and 15 showed higher diversity than DYS393 (diversity=0.66; Table 2). The discrimination of haplotypes that are not distinguished by the commonly used markers is a particularly useful property. As reported [10], eight pairs of YCC individuals carry haplotypes that are identical when the standard minimal set of Y-STRs is used. Two of these are from different populations (Mbuti Pygmy/Bantu speaker; English/German) and these were distinguished by seven and nine of the new loci, respectively. The other six pairs are from isolated populations, and these were distinguished by 2, 1, 1, 0, 0 and 0, respectively, of the new markers (Table S4). Although a total of 15 loci contribute to this increased discrimination, all of the five distinguishable haplotypes could be separated using just two of the most variable loci, DYS570 and DYS576.

Discussion

We have investigated the properties of 52 new Y-STRs in a diverse worldwide set of males. We found that two of the

Table 2 Variation of 50 new Y-STR loci in the YCC panel

Locus	Mean repeat count	Number of alleles	Diversity	Variance
DYS481	23.3	11	0.90	7.87
DYS570	17.6	11	0.86	3.89
DYS576	17.3	7	0.82	2.16
DYS643	11.1	9	0.82	2.33
DYS485	15.9	7	0.78	1.92
DYF406S1	10.6	7	0.75	1.26
DYS522	11.2	6	0.74	1.03
DYS589	12.3	6	0.73	1.07
DYS533	11.2	6	0.72	1.03
DYS549	12.1	5	0.72	0.92
DYS505	12.1	6	0.71	1.00
DYS508	11.4	8	0.71	1.44
DYS525	9.9	7	0.71	0.95
DYS531	11.0	4	0.69	0.39
DYS556	11.7	6	0.67	1.03
DYS572	10.3	5	0.64	0.61
DYS565	11.5	6	0.64	0.71
DYS594	9.5	7	0.63	0.87
DYS540	11.4	4	0.62	0.52
DYS487	13.2	6	0.62	1.38
DYS511	10.9	5	0.62	0.73
DYS573	9.9	4	0.62	0.73
DYS617	12.4	5	0.62	1.06
DYS495	15.3	5	0.61	0.76
DYS567	10.4	6	0.61	0.72
DYS497	14.2	4	0.60	0.51
DYS488	13.4	6	0.58	0.91
DYS492	11.6	5	0.56	0.45
DYS490	12.9	7	0.56	2.43
DYS568	11.1	7	0.56	0.92
DYS636	11.3	5	0.56	0.49
DYS537	10.9	4	0.53	0.43
DYS618	11.8	4	0.51	0.37
DYS638	10.9	4	0.49	0.34
DYS491	12.1	5	0.46	0.40
DYS578	8.1	4	0.41	0.40
DYS640	11.2	3	0.39	0.39
DYS476	11.2	4	0.36	0.29
DYS494	9.0	4	0.35	0.28
DYS641	9.6	4	0.33	0.93
DYS554	9.0	5	0.31	0.36
DYS575	10.0	4	0.25	0.22
DYS583	8.0	3	0.22	0.12
DYS590	7.9	3	0.22	0.23
DYS480	7.9	3	0.22	0.16
DYS569	11.0	3	0.18	0.11
DYS530	9.1	2	0.17	0.09
DYS580	9.1	4	0.11	0.19
DYS472	8.0	2	0.08	0.04
DYS579	9.0	2	0.05	0.89

Loci are ordered according to their diversity.

Y-STRs were multicopy and thus not well suited to some applications and that the remaining 50 loci differed substantially in their properties. Our measurements of allele numbers, diversity and variance were overall consistent with the previous report [5]; correlation coefficients (R^2 values) were 0.47, 0.58 and 0.67, respectively, but differed for some individual loci. The most variable Y-STR, in all respects, was DYS481, and this was not previously considered in detail because sequence data were not available before. Several other loci (e.g., DYS570, DYS576 and DYS643) may be particularly useful for increasing discrimination in forensic work, and the simple structure and mutational properties of this set make them the markers of choice for many population genetic studies. This is illustrated by considering the correlation between mean repeat count and variance in repeat number of the 50 simple loci: it was far higher ($R^2=0.67$) than the value reported for complex Y-STRs ($R^2=0.34$, [5]), suggesting that the simple STRs have simpler mutational mechanisms and may lead to more precise dates of lineages. The data in Table 2 and Table S3 now provide a basis for choosing the best simple loci and assembling them into a high-level multiplex reaction for more extensive population screening.

Acknowledgments We thank the original sample donors and Mike Hammer and Nathan Ellis for providing the YCC DNA samples, Peter de Knijff for YCC haplotype data, John Butler for the AutoDimer software before publication, Elizabeth Huckle for sequencing and Denise Carvalho-Silva for help during the course of this work. We particularly thank Manfred Kayser for helpful comments and corrections. S-KL was supported by a Korean Government short-term fellowship for overseas study, EJP by a grant from the Arts and

Humanities Research Council and the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409 and YX and CT-S by The Wellcome Trust.

References

1. Gusmão L, Butler JM, Carracedo A et al (2005) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int J Legal Med* 119:1–10
2. Jobling MA, Hurler ME, Tyler-Smith C (2004) *Human evolutionary genetics*. Garland Science, New York and Abingdon
3. Roewer L, Krawczak M, Willuweit S et al (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int* 118:106–113
4. Xue Y, Zerjal T, Bao W et al (2005) Recent spread of a Y-chromosomal lineage in northern China and Mongolia. *Am J Hum Genet* 77:1112–1116
5. Kayser M, Kittler R, Erler A et al (2004) A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet* 74:1183–1197
6. Butler JM, Decker AE, Vallone PM, Kline MC (2006) Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci Int* 156:250–260
7. Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348
8. Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20:1004–1006, 1008–1010
9. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* 37:226–231
10. Redd AJ, Agellon AB, Kearney VA et al (2002) Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int* 130:97–111