Available online at www.sciencedirect.com
SCIENCE DIRECT
Forensic Science International 161 (2006) 1–7
Forensic Science International
www.elsevier.com/locate/forsciint

26-Locus Y-STR typing in a Bhutanese population sample

Emma J. Parkin [a], Thirsa Kraayenbrink [b], George L. van Driem [c], Karma Tshering of Gaselô [b], Peter de Knijff [b], Mark A. Jobling [a,*]

[a] *Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK*
[b] *MGC-Department of Human and Clinical Genetics, Leiden University Medical Centre, The Netherlands*
[c] *Himalayan Languages Project, Leiden University, The Netherlands*

Received 10 August 2005; received in revised form 9 October 2005; accepted 9 October 2005
Available online 14 November 2005

Abstract

26 Y chromosome short tandem repeat (STR) loci were amplified in a sample of 856 unrelated males from Bhutan, using two multiplex polymerase chain reaction (PCR) assays. The first multiplex is the Y-STR 20plex described by Butler et al. [J.M. Butler, R. Schoske, P.M. Vallone, M.C. Kline, A.J. Redd, M.F. Hammer, A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers, Forensic Sci. Int. 129 (2002) 10–24], and the second is a novel (but overlapping) 14plex that targets six additional Y-STRs (DYS425, DYS434, DYS435, DYS436, DYS461, DYS462) and also amplifies the amelogenin locus. The 26-loci give a discriminating power of 0.9957, though even at this resolution one haplotype occurs 24 times. We identify novel alleles at five loci and microvariants at a further three, which were characterised by sequencing. Extended (11-locus) haplotypes for these samples have been submitted to the Y-STR Haplotype Reference Database (YHRD).

Keywords: Y chromosome; STRs; Microsatellites; Multiplex; Bhutan

1. Introduction

Y-chromosomal short tandem repeats (STRs) are becoming increasingly widely applied in forensic analysis because of their male-specificity and the informativeness of the haplotypes they define. Enough markers are available to provide useful discriminating power in individual identification work, and the geographical differentiation of Y haplotypes [1] means that information about the population-of-origin of the donor of a given Y chromosome can often be deduced.

Because of the strong geographical structure of Y-chromosomal haplotype variation, the usefulness of Y haplotypes in forensic casework depends on the availability of accurate and extensive population databases. The online and publicly available Y-STR Haplotype Reference Database (YHRD) is an excellent such resource, but currently has a strong European bias. Here, we describe alleles at 26 Y-STRs, and properties of the haplotypes they define, in a previously unstudied population, that of Bhutan in the eastern Himalayas. Eleven-locus haplotypes have been submitted to the YHRD, and full data are available from the authors on request. Our report follows guidelines for the publication of population data [2].

Sampling of 856 Bhutanese males was undertaken as part of a larger collaborative project [3] examining genetic diversity in populations of the Himalayas within the context of their extraordinary degree of linguistic diversity [4]. Here, we describe our initial findings with Y-STRs, treating the Bhutanese sample as a single population; relationships between sub-populations and linguistic sub-groups, together with data on other genetic markers, will be described in future publications. The samples represent 19 distinct ethnolinguistic

* Corresponding author. Tel.: +44 116 252 3427; fax: +44 116 252 3378.
E-mail address: maj4@leicester.ac.uk (M.A. Jobling).

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2005.10.008

groups, all speaking Tibeto–Burman languages, and widely distributed throughout Bhutan.

Our choice of Y-STRs was governed by three factors: comparability with a broad range of published data sets on other populations, forensic utility and efficiency of typing. These requirements were met in two multiplexes, one of which has been previously described [5] and amplifies 20 Y-STRs, and the other which is novel, and targets 13 Y-STRs, seven of which are shared with the 20plex, plus amelogenin.

2. Materials and methods

2.1. DNA samples

Eight hundred and fifty-six Bhutanese males provided blood samples with informed consent, and DNA was extracted using Autopure LS® from Gentra Systems, according to the manufacturer's specifications. DNA samples from collections of the authors, including Y Chromosome Consortium (YCC) cell lines [6] were used as haplotype reference materials.

2.2. Y-STR multiplexes

2.2.1. 20plex

The multiplex design and the 34 primer sequences required have been described previously [5]. This 20plex amplifies the following loci and includes all 11 Y-STR markers in the European "extended haplotype" (given in bold; http://www.yhrd.org/): **DYS19**, **DYS385a/b**, DYS388, **DYS389I**, **DYS389II**, **DYS390**, **DYS391**, **DYS392**, **DYS393**, DYS426, DYS437, **DYS438**, **DYS439**, DYS447, DYS448, DYS460, YCAIIa/b and Y-GATA-H4.1.

2.2.2. 14plex

In summary, this multiplex includes the amelogenin sex-test [7], and targets 13 Y-STRs. Six of these, namely DYS19, DYS385a/b, DYS392, DYS438, DYS447 and DYS448 are also typed by the 20plex using the same primers; this redundancy provides a cross-check for sample identity. Primers for five of the remaining six Y-STRs and for amelogenin are as described by Bosch et al. [8] and those for DYS425 are as described by Jobling et al. [9]. The design of this novel multiplex and the 5′ end-labelled fluorescent dye of each forward primer are illustrated in Fig. 1. In order to accommodate DYS19 [5] within the design of the 14plex it was necessary to alter its 5′ fluorescent label from NED™ (yellow) to 6-FAM™ (blue).

2.3. Polymerase chain reaction (PCR) amplification conditions

PCR amplifications were performed using 5 ng template DNA in a total reaction volume of 10 μl containing 1 unit of AmpliTaq Gold DNA polymerase (Applied Biosystems),

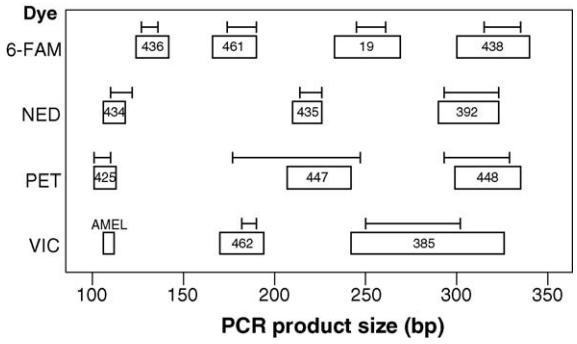

Fig. 1. Schematic representation of allele size ranges and dyes associated with STR loci within the 14plex. For each locus, the box indicates the size range of previously observed alleles in published studies ([14] and STRBase) and the bar above indicates the observed allele size range in this study. 'AMEL', amelogenin.

10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.75 mM $MgCl_2$, 300 μM dNTPs and 5% (v/v) glycerol. For the 20plex PCR, primer concentrations were according to Butler et al. [5]; Table 1 lists primer concentrations for the 14plex. PCR cycling conditions were: 95 °C for 10 min, 28 cycles of 94 °C for 1 min, 55 °C for 1 min and 65 °C for 1 min and final extension of 65 °C for 45 min.

To ensure that the resultant PCR amplicons are A-tailed (thereby avoiding the split peak phenomenon when visualised), a 4 μl reaction mix incorporating 0.25 U *Taq* polymerase (ABgene) and a 1× PCR buffer system [10] was added to the PCR products prior to incubation for a further 45 min at 72 °C.

2.4. Detection and analysis of PCR products

The A-tailed PCR products were diluted (1 in 5) and 2 μl of this dilution mixed with 10 μl Hi-Di™ formamide and 0.15 μl GS500 LIZ size standard (Applied Biosystems). After incubation at 95 °C for 3 min, followed by 5 min on ice, the samples were loaded onto an ABI Prism® 3100 Genetic Analyzer 16-capillary array system (Applied Biosystems). The G5 matrix filter was used to detect the five dyes 6-FAM™ (blue), VIC™ (green), NED™ (yellow), PET™ (red) and LIZ™ (orange). The samples were injected for 22 s at 3000 V. Separations were performed at 15,000 V for 25 min with a run temperature of 60 °C using the 3100 POP™-4 sieving polymer (Applied Biosystems) and 1× genetic analyzer buffer with EDTA, run through a 36 cm capillary array (Applied Biosystems).

Data were analysed on a PC with Genemapper v. 3.0 (Applied Biosystems). Bins of 1.5 bp were created (see Fig. 1) in order to discriminate allelic forms and the allele calls were assigned using multiplex PCR products derived from reference DNA samples. For the 20plex, YCC cell line DNAs [6] were used as reference DNAs, as their full haplotypes are provided in the literature [5]. Allele designation for the 14plex relied on a combination of YCC samples

Table 1
Characteristics of PCR multiplexes used in this study

STR locus	20plex		14plex		Repeat motif	Reference
	Dye	Final primer concentration (μM)	Dye	Final primer concentration (μM)		
20plex only						
DYS388	NED	1.80			ATT	[5]
DYS389I/II	6-FAM	1.30			TCTG/A	[5]
DYS390	VIC	0.20			TCTG/A	[5]
DYS391	6-FAM	0.20			TCTA	[5]
DYS393	VIC	0.40			AGAT	[5]
DYS426	VIC	0.20			GTT	[5]
DYS437	6-FAM	0.20			TCTA/G	[5]
DYS439	6-FAM	0.20			GATA	[5]
DYS460	NED	0.70			ATAG	[5]
Y-GATA-H4.1	NED	0.40			AGAT	[5]
YCAIIa/b	VIC	0.90			CA	[5]
Shared loci						
DYS19	NED	1.80	6-FAM	0.45	TAGA	[5]
DYS385	VIC	0.60	VIC	0.11	GAAA	[5]
DYS392	NED	1.80	NED	0.23	TAT	[5]
DYS438	6-FAM	0.20	6-FAM	0.23	TTTTC	[5]
DYS447	PET	0.70	PET	0.70	TAATA	[5]
DYS448	PET	1.10	PET	0.23	AGAGAT	[5]
14plex only						
DYS425			PET	0.23	TGT	[9]
DYS434			NED	0.10	CTAT/TAAT	[8]
DYS435			NED	0.09	TGGA	[8]
DYS436			6-FAM	0.08	GTT	[8]
DYS461			6-FAM	0.11	T/CAGA	[8]
DYS462			VIC	0.08	TATG	[8]
AMEL			VIC	0.05	n/a	[8]

The forward primers were 5′-fluoresecently labelled with the dyes indicated.

and DNAs from collections of the authors. Allele-sizing data, provided by the internal-sizing standard, confirmed the allelic identities based on expectations derived from published sequence data.

2.5. Quality control

In addition to regularly typing the reference DNAs to confirm that the typing remained consistently accurate, validation of the PCR method and allele-calling was carried out by blind retyping of 60 DNA samples; all alleles called in the original typings were correctly called in the repeat typings. The quality control exercise for submission of data to the YHRD was also passed.

2.6. Y-STR nomenclature

Allele nomenclature was according to Butler et al. [5] and Bosch et al. [8], with the exception of DYS439, DYS448 and Y-GATA-H4.1, where it was necessary to change nomenclature for compatibility with ISFG recommendations [11]. Compared to Butler et al. [5], seven repeats were subtracted from DYS439, three subtracted from DYS448 and eight added to Y-GATA-H4.1. Note that Y-GATA-H4.1 was previously referred to as Y-GATA-H4 [5], a name now used to refer to a larger amplicon including additional repeats [11]; in addition, the repeat motif is now (AGAT), rather than (TAGA). For DYS425, the repeat motif was taken to be (TGT) [12], and alleles in three reference chromosomes [6] defined by sequencing: YCC15 (allele 13), YCC24 (allele 12) and YCC43 (allele 12).

2.7. Calculations

Gene diversity and haplotype diversity were calculated using Arlequin [13].

3. Results and discussion

3.1. Diversity of alleles

Tables 2 and 3 show the allele frequency distributions for all the Y-STRs studied. As in a previous study [8], DYS385

Table 2
Frequencies of alleles at 22 of the 26 Y-STRs

Allele	19	388	390	391	392	393	425	426	434	435	436	437	438	439	447	448	460	461	462	389I	389II–I	H4.1
7					0.030												0.019					
8				0.001										0.001			0.048					
9				0.119		0.002							0.046				0.472					
10		0.600		0.738	0.096				0.098	0.001	0.005		0.167	0.110			0.239	0.001				
11		0.004		0.139	0.140	0.020	0.001	0.857	0.863	0.977			0.716	0.362			0.194	0.131	0.121	0.002		
12		0.345		0.002	0.020	0.717	0.901	0.143	0.028	0.019	0.952		0.070	0.339			0.028	0.729	0.783	0.558		
13	0.041	0.047			0.035	0.200	0.042		0.011	0.004	0.042		0.001	0.171				0.131	0.096	0.248	0.011	
14	0.409	0.001			0.650	0.054						0.387		0.016				0.008		0.186		
15	0.473	0.002			0.025	0.007						0.541								0.006	0.047	
16	0.050				0.001							0.070			0.001	0.001					0.671	
17	0.023	0.001			0.004							0.002				0.036					0.202	
18																0.030					0.056	0.055
19																0.244					0.014	0.218
20															0.001	0.582						0.574
21															0.023	0.098						0.150
22			0.025												0.016	0.007						0.004
23			0.595												0.555							
24			0.252												0.113							
25			0.114												0.050							
26			0.013												0.085							
27			0.001												0.039							
28															0.050							
29															0.026							
30															0.004							
Microvariants, duplicated and null alleles																						
11.1[a]											0.001											
20.4																0.001						
21.4															0.005							
22.4															0.032							
10–11							0.010							0.001								
11–12							0.019															
12–13							0.021															
16–17	0.001																					
Null	0.002						0.006															
h[b]	0.604	0.518	0.569	0.421	0.546	0.442	0.187	0.244	0.244	0.046	0.092	0.553	0.452	0.713	0.663	0.590	0.679	0.434	0.363	0.592	0.504	0.598

[a] The designation 11.1 is the repeat number based on sequencing (see Fig. 2a); however, the amplicon size suggests a 10.2 repeat allele.
[b] Calculation of gene diversity, h, excludes null alleles and duplications.

Table 3
Frequencies of genotypes at DYS385 and YCAII

Genotype	DYS385	YCAII
9–16	0.002	
9–17	0.002	
9–18	0.001	
10–12	0.002	
11–11	0.005	
11–12	0.028	
11–13	0.013	
11–14	0.048	
11–15	0.005	
11–16	0.001	
11–17	0.002	
11–18	0.002	
11–19	0.001	
11–20	0.002	
12–12	0.020	
12–13	0.011	
12–14	0.009	
12–16	0.005	
12–17	0.009	
12–18	0.015	
12–19	0.009	
12–20	0.002	
13–13	0.005	
13–16	0.030	
13–17	0.049	
13–18	0.224	
13–19	0.099	
13–20	0.035	
13–21	0.033	
13–22	0.008	
14–14	0.001	
14–15	0.004	
14–16	0.009	
14–17	0.033	
14–18	0.036	
14–19	0.037	
14–20	0.018	
14–21	0.005	
15–16	0.037	0.008
15–17	0.004	0.018
15–18	0.005	0.001
15–19	0.002	
15–20	0.018	
15–21	0.011	
15–22	0.002	
16–16	0.004	
16–17	0.029	0.001
16–18	0.023	
16–19	0.001	0.006
16–21	0.002	
17–17	0.021	0.057
17–18	0.010	0.058
17–19	0.001	0.678
17–20		0.055
17–21		0.078
18–18	0.002	0.008
18–19		0.014
19–19		0.005
19–20		0.002
19–21		0.001
20–20		0.007
20–21		0.002
13–17–18	0.001	
13–18–19	0.001	
13–18–20	0.001	
12.2–20	0.001	
15.2–20	0.001	
h[a]	0.921	0.524

[a] Calculation of gene diversity, h, excludes duplicated alleles.

(when considered as a genotype—Table 3) is the most diverse marker within the Y-STR set, with a gene diversity (h) of 0.921 and the most polymorphic single-locus marker is DYS439 ($h = 0.713$). The marker DYS435 is of markedly lower diversity than the rest, with 97.7% of chromosomes carrying allele 11. There seems little reason to type this marker in forensic contexts, though as a slow-mutating STR it still has potential in evolutionary studies, and for our purposes allows comparison with other datasets.

Previously unreported alleles (defined with reference to Butler [14] and STRBase (http://www.cstl.nist.gov/biotech/strbase/index.htm)) were found at five loci, as follows: with those characterised by sequencing indicated by asterisks: DYS434 (allele 13), DYS439 (allele 8*), DYS447 (alleles 16*, 20, 21 and 30*), DYS448 (allele 16) and Y-GATA-H4.1 (allele 22).

'Null' alleles or multiple peaks were reproducibly obtained at a number of loci. For DYS19, two individuals carried null alleles, while one carried both alleles 16 and 17; for DYS385, three individuals each carried three alleles (see Table 3); and for DYS439, one individual carried both alleles 10 and 11. Multiple peaks in these cases are most simply explained by locus duplication. The trinucleotide STR DYS425 is a more complex case, since it is one locus within a multi-locus STR, amplified using a locus-specific forward primer targeted at a single specific base. It has been shown that gene conversion (non-reciprocal transfer of sequence information) occurs frequently between repeated sequences on the Y chromosome [15,16], so the occurrence of null or multiple peak cases at DYS425 may be better explained by a conversion event transferring a primer-binding site between copies, rather than conventional deletion or duplication processes.

Microvariants (partial alleles) were present at four STR loci (Tables 2 and 3). Those at DYS385 were not investigated further, but those at DYS436, DYS447 and DYS448 were analysed by sequencing, the results of which are summarised in Fig. 2. The microvariant observed at the trinucleotide locus DYS436 involves a 4 bp deletion which straddles the

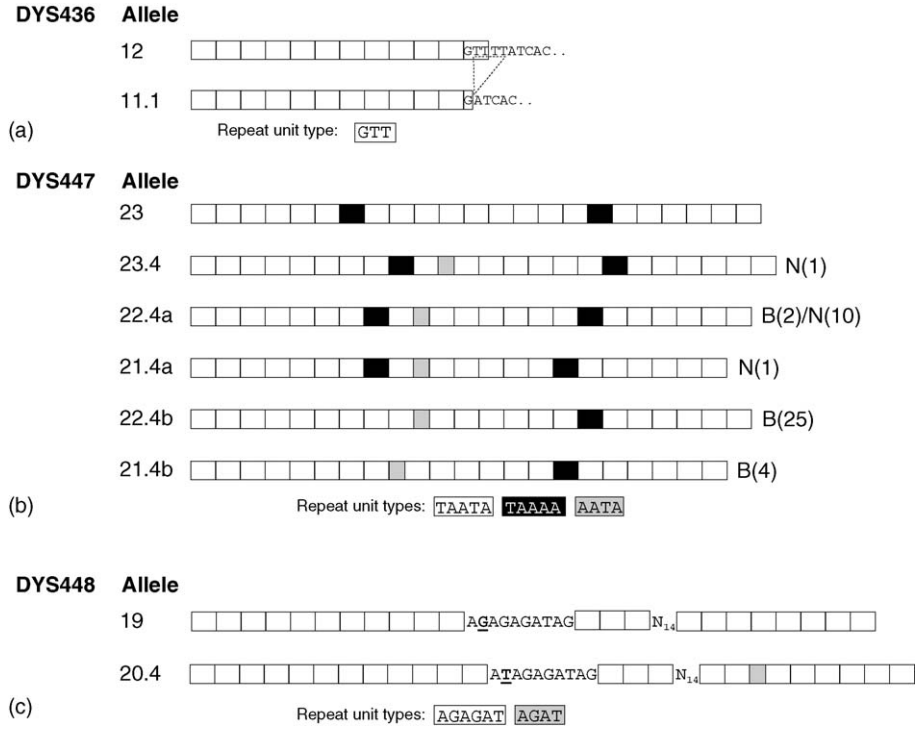

Fig. 2. Sequence structures of microvariants. In each case, the upper allele is that seen in the genome database sequence. (a) DYS436: microvariant observed in a single individual, (b) DYS447: structures of five different alleles are shown schematically; pairs marked a and b are isoalleles. Alleles marked 'B' are found the in the Bhutanese samples described here, while alleles marked 'N' are found in a large Nepalese population sample to be described elsewhere, and are included here for completeness. Incidences of these alleles are given in parentheses and (c) DYS448: microvariant observed in a single individual. Base substitutional variant in bold and underlined. Note that the internal block of three repeats is not included in the overall repeat count for this locus.

repeat array and the 3′ flanking sequence; the length of the PCR product therefore suggests allele 10.2, but sequencing shows that it actually contains 11.1 repeat units. Microvariants at DYS447 result from a deletion of 1 bp within the pentanucleotide repeat array; sequencing reveals isoallelism, in that alleles designated 21.4 and 22.4 each exist in two forms with different internal structures. The DYS448 microvariant includes a G–T transversion in the 10-bp segment that separates the first two blocks of hexanucleotide repeats.

3.2. Diversity of haplotypes

For simplicity, haplotype diversity (equivalent to power of discrimination, PD) was calculated with the omission of chromosomes lacking null alleles and duplications, providing a sample size of 802. For the full set of 26 Y-STRs, there are 299 unique haplotypes (37.3%) and PD is 0.9957. The corresponding values for the 20plex [5], extended (11-locus) haplotype and minimal (9-locus) haplotype are shown in Fig. 3. Notably, the addition of a further six STRs to the 20plex did not significantly improve its power of discrimination.

Fig. 3 also illustrates the distribution of non-unique haplotypes for the full set of 26 Y-STRs, the 20plex [5] and the extended and minimal haplotypes of the YHRD. High-frequency haplotypes are of particular interest, and in the 802 males we observe nine instances of 26-locus haplotypes that are present in 10 or more individuals (Fig. 3a), including one that is observed 24 times. Of these nine common haplotypes, seven (including the most common haplotype) are predominantly (≥90%) restricted to particular geographically defined sub-populations. A future publication will explore the significance of these high-frequency haplotypes more fully.

3.3. Concluding remarks

Y-STR typing at the resolution carried out here provides considerable discriminating power for forensic analysis and the haplotypes defined are also highly informative in population genetic studies. The Bhutanese sample set is currently being analysed using Y-chromosomal binary markers and the combination of STR haplotypes and binary haplogroups will provide the first insights into the peopling of this region of the world.

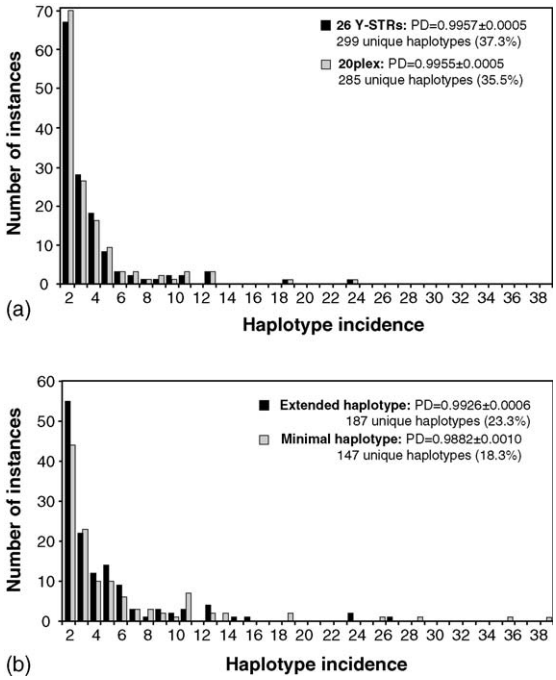

Fig. 3. Haplotype diversity for (a) all 26 STRs and the 20plex and (b) the extended and minimal haplotypes. Histograms show the frequency distributions of haplotypes present more than once in the dataset.

Acknowledgements

We thank all DNA donors. This research was conducted as part of the longstanding bilateral cooperation between the Royal Government of Bhutan and Leiden University, and sample collection was carried out in accordance with the Memorandum of Understanding between the Dzongkha Development Authority (DDA) and Leiden University. As part of the European Science Foundation EUROCORES Programme OMLL, this work was supported by the Arts and Humanities Research Council and the EC Sixth Framework Programme under Contract no. ERAS-CT-2003-980409. T.K., G.L.vD., K.T.G. and P.dK. were supported by funds from the Netherlands Organisation for Scientific Research (NWO grant no. 231-70-001). M.A.J. was supported by a Wellcome Trust Senior Fellowship in Basic Biomedical Science (grant no. 057559).

References

[1] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, Nat. Rev. Genet. 4 (2003) 598–612.

[2] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, Forensic Sci. Int. 110 (2000) 3–5.

[3] T. Kraayenbrink, P. de Knijff, G.L. van Driem, J.R.M.L. Opgenort, M.A. Jobling, E.J. Parkin, C. Tyler-Smith, D.R. Carvalho-Silva, K. Tshering, G. Barbujani, I. Dupanloup, G. Bertorelle, N.M. Tuladhar, Language and genes of the Greater Himalayan region, in: R. Klein (Ed.), The Origin of Man, Language and Languages, European Science Foundation, Strasbourg, in press.

[4] G.L. van Driem, Languages of The Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region containing an Introduction to the Symbiotic Theory of Language, 2 vols., Brill, Leiden, 2001.

[5] J.M. Butler, R. Schoske, P.M. Vallone, M.C. Kline, A.J. Redd, M.F. Hammer, A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers, Forensic Sci. Int. 129 (2002) 10–24.

[6] Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, Genome Res. 12 (2002) 339–348.

[7] K.M. Sullivan, A. Mannucci, C.P. Kimpton, P. Gill, A rapid and quantitative DNA sex test—fluorescence-based PCR analysis of X–Y homologous gene amelogenin, Biotechniques 15 (1993) 636.

[8] E. Bosch, A.C. Lee, F. Calafell, E. Arroyo, P. Henneman, P. de Knijff, M.A. Jobling, High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions, Forensic Sci. Int. 125 (2002) 42–51.

[9] M.A. Jobling, V. Samara, A. Pandya, N. Fretwell, B. Bernasconi, R.J. Mitchell, T. Gerelsaikhan, B. Dashnyam, A. Sajantila, P.J. Salo, Y. Nakahori, C.M. Disteche, K. Thangaraj, L. Singh, M.H. Crawford, C. Tyler-Smith, Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males, Hum. Mol. Genet. 5 (1996) 1767–1775.

[10] A.J. Jeffreys, R. Neumann, V. Wilson, Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis, Cell 60 (1990) 473–485.

[11] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider, DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, Forensic Sci. Int. 157 (2006) 187–197.

[12] D.L. Daniels, A.M. Hall, J. Ballantyne, SWGDAM developmental validation of a 19-locus Y-STR system for forensic casework, J. Forensic Sci. 49 (2004) JFS2003134.

[13] S. Schneider, D. Roessli, L. Excoffier, Arlequin Version 2.0: A Software for Population Genetics Data Analysis, second ed., Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland, 2000.

[14] J.M. Butler, Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis, Forensic Sci. Rev. 15 (2003) 91–111.

[15] S. Rozen, H. Skaletsky, J.D. Marszalek, P.J. Minx, H.S. Cordum, R.H. Waterston, R.K. Wilson, D.C. Page, Abundant gene conversion between arms of massive palindromes in human and ape Y chromosomes, Nature 423 (2003) 873–876.

[16] E. Bosch, M.E. Hurles, A. Navarro, M.A. Jobling, Dynamics of a human interparalog gene conversion hotspot, Genome Res. 14 (2004) 835–844.