Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International 170 (2007) 68–72
Forensic Science International
www.elsevier.com/locate/forsciint

Announcement of Population Data

Allele frequency distribution for 21 autosomal STR loci in Bhutan

Thirsa Kraaijenbrink [a], George L. van Driem [b], Karma Tshering of Gaselô [b], Peter de Knijff [a,*]

[a] *Department of Human Genetics, Postzone S5-P, Leiden University Medical Centre, P.O. Box 9600, 2300 RC, Leiden, The Netherlands*
[b] *Himalayan Languages Project, FBJ Kuiperinstituut, Leiden University, P.N. van Eyckhof 3, 2311 BV Leiden, The Netherlands*

Received 1 March 2006; received in revised form 5 April 2006; accepted 11 April 2006
Available online 24 May 2006

Abstract

We studied the allele frequency distribution of 21 autosomal STR loci contained in the AmpF*l*STR® Identifiler™ (Applied Biosystems), the Powerplex®16 (Promega) and the FFFL (Promega) multiplex PCR kits among 936 individuals from the Royal Kingdom of Bhutan. As such these are the first published autosomal DNA results from this country.

Keywords: Bhutan; Identifiler; Powerplex®16; FFFL

Population information: The Royal Kingdom of Bhutan lies just south of the highest peaks in the eastern section of the Himalayan mountain range, which runs from Pakistan in the west to north-eastern India (Assam) in the east. As such, the Himalayas are not only the highest land barrier on earth, they also form a distinct barrier between Tibeto-Burman, which is predominantly spoken north-east of the Himalayas, and Indo-European, which is spoken south of the Himalayas [1]. In Bhutan, nearly all people speak a language belonging to the Tibeto-Burman language family. There are many distinct isolated languages within Bhutan, many even not completely characterized. It has been suggested that in the past humans speaking a Tibeto-Burman language either have crossed the Himalayas, or have navigated around the highest parts of the Himalayas before settling in Bhutan. In order to be able to reconstruct possible migration routes connecting human populations North and South of the Himalayas, we collected blood samples from 761 males and 175 females from Bhutan, belonging to 17 different language groups defined by van Driem [1] (see Fig. 1). Here we describe the first results of this study in the form of the allele frequency distributions and summary statistics of 21 different autosomal STR loci from the combined group of 936 Bhutanese individuals. To our knowledge, this is the first genetic study on autosomal STR loci from the Royal Kingdom of Bhutan.

DNA extraction: DNA was extracted from whole blood, using the Autopure LS® from Gentra Systems, according to the manufacturer's specifications.

PCR and genotyping: Identifiler™, Powerplex®16 and FFFL® PCR reactions were performed according to the manufacturers' specifications. PCR products were analysed using an ABI 3100 automated DNA sequencer and the Genemapper®ID software.

Statistical analysis of data: Allele frequencies and heterozygosity values (observed and expected) were calculated using GenePop version 3.4 [2], exact Hardy–Weinberg *p*-values and PIC values were calculated using PowerMarker [3]. The Excel PowerStats spreadsheet [4] was used to calculate power of discrimination and power of exclusion. We used the exact test of population differentiation included in Arlequin Ver. 2000 [5] to compare differences in allele frequencies among different populations.

Results: Allele frequencies and summary statistics are presented in Table 1. We identified a number of new alleles (i.e. not yet reported in the NIST Short Tandem Repeat DNA Internet DataBase [6]). These are shown underlined in Table 1. Using the PowerMarker software, markers D16S539, D8S1179, D21S11, FGA, TPOX and LPL were found to have significant HWE *p*-values (i.e. below the Bonferroni corrected *p*-value threshold of 0.0023). Similar results were obtained by using GenePop. Further, GenePop analysis showed significant heterozygote deficiency for FGA. Subsequently, in order to test for the presence of undetected null alleles as a cause for heterozygote deficiency, the data were analysed using the

* Corresponding author. Tel.: +31 71 526 9540; fax: +31 71 526 8278.
E-mail address: knijff@lumc.nl (P. de Knijff).

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2006.04.006

Table 1
Allele frequencies for 21 STR loci in 936 Bhutanese samples

	D3S1358	VWA	D16S539	D8S1179	D21S11	D18S51	TH01	FGA	D2S1338	D19S433	PENTA E	D5S818	D13S317	D7S820	CSF1PO	PENTA D	TPOX	F13A01	F13B	FES	LPL
3.2																		0.2297			
4																		0.1934			
5											0.0256							0.1170			
6							0.0887									0.0032		0.4348	0.0027		
7							0.2607				0.0032	0.0069	0.0005	0.0021	0.0005	0.0214	0.0027	0.0123	0.0005		
8			0.0422				0.0897				0.0011		0.1725	0.2126	0.0005	0.0689	0.5395	0.0059	0.1223	0.0011	
8.2																0.0021					
9			0.2559				0.4690				0.0144	0.0721	0.1234	0.0769	0.0497	0.2174	0.1501	0.0032	0.1966	0.0011	0.0016
9.3							0.0860														
10			0.1106	0.0668		0.0005	0.0059			0.0011	0.0171	0.1784	0.1640	0.1966	0.2009	0.1400	0.0069	0.0005	0.6752	0.0246	0.6442
11			0.2601	0.0347		0.0043				0.0048	0.1036	0.3536	0.2975	0.2885	0.2821	0.2441	0.2585		0.0027	0.5214	0.1282
11.2										0.0005											
12			0.2393	0.1074		0.0208				0.0684	0.1560	0.2105	0.1715	0.1859	0.3873	0.1501	0.0411			0.2788	0.2147
12.2										0.0166											
13	0.0005		0.0812	0.2698		0.2441				0.1976	0.0636	0.1651	0.0561	0.0363	0.0759	0.0919	0.0011			0.1608	0.0101
13.2										0.0214											
14	0.0342	0.1763	0.0069	0.1795		0.1955				0.2596	0.0929	0.0123	0.0139	0.0011	0.0032	0.0507		0.0005		0.0123	0.0011
14.2										0.1303											
15	0.3312	0.0171	0.0037	0.2131		0.1928				0.0956	0.1154	0.0011	0.0005			0.0085		0.0016			
15.2										0.1544											
16	0.3665	0.2270		0.1106		0.0743			0.0118	0.0160	0.1293					0.0011		0.0005			
16.2										0.0272											
17	0.2158	0.2591		0.0160		0.0775		0.0005	0.0764	0.0064	0.0534					0.0005		0.0005			
18	0.0491	0.2051		0.0021		0.0427		0.0497	0.1202		0.0705										
19	0.0027	0.0999				0.0732		0.1165	0.1880		0.0443										
19.3									0.0005												
20		0.0144				0.0433		0.0262	0.1213		0.0438										
20.2								0.0118													
21		0.0011				0.0187		0.0545	0.0363		0.0395										
21.2								0.0134													
22						0.0101		0.1506	0.0192		0.0128										
22.2								0.0144													
23						0.0016		0.1741	0.2163		0.0112										
23.2								0.0240													
24						0.0005		0.1683	0.1405												
24.2								0.0235													
25								0.0721	0.0620		0.0005										
25.2								0.0053													
26								0.0449	0.0037		0.0016										
26.2								0.0053													
27					0.0112			0.0198	0.0037												
27.2								0.0005													
28					0.0321			0.0187													
28.2					0.0556																
29					0.2350			0.0032													
29.2					0.0091																
30					0.2286			0.0027													
30.2					0.0379																

Table 1 (*Continued*)

	D3S1358	VWA	D16S539	D8S1179	D21S11	D18S51	TH01	FGA	D2S1338	D19S433	PENTA E	D5S818	D13S317	D7S820	CSF1PO	PENTA D	TPOX	F13A01	F13B	FES	LPL
31					0.1111																
31.2					0.0417																
32					0.0288																
32.2					0.1485																
33					0.0011																
33.2					0.0577																
34.2					0.0016																
N	936	936	936	936	936	936	936	936	936	936	936	936	936	936	936	936	936	936	936	936	936
OH	0.6998	0.7735	0.7447	0.7436	0.8002	0.8002	0.6613	0.8376	0.8205	0.8269	0.8761	0.7415	0.7821	0.7511	0.6902	0.8056	0.5598	0.6891	0.4733	0.5983	0.4712
EH	0.7062	0.7981	0.7893	0.8203	0.8469	0.8440	0.6891	0.8898	0.8579	0.8375	0.9080	0.7666	0.8073	0.7916	0.7223	0.8351	0.6181	0.7073	0.4907	0.6241	0.5226
HWE	0.0866	0.1529	**0.0020**	**0.0001**	**0.0001**	0.0354	0.0026	**0.0000**	0.0322	0.0422	0.2892	0.0577	0.4321	0.0060	0.6507	0.2430	**0.0000**	0.7743	0.6935	0.3796	**0.0003**
PIC	0.6516	0.7668	0.7567	0.7962	0.8293	0.8255	0.6450	0.8795	0.8414	0.8176	0.9002	0.7309	0.7801	0.7592	0.6750	0.8142	0.5614	0.6611	0.4404	0.5629	0.4688
PD	0.8561	0.9291	0.9242	0.9458	0.9597	0.9590	0.8535	0.9770	0.9640	0.9532	0.9836	0.9099	0.9367	0.9255	0.8799	0.9512	0.7966	0.8699	0.6905	0.7998	0.7148
PE	0.4279	0.5508	0.5007	0.4989	0.5994	0.5994	0.3710	0.6706	0.6376	0.6499	0.7468	0.4953	0.5662	0.5116	0.4132	0.6094	0.2454	0.4116	0.1652	0.2888	0.1635

OH, observed heterozygosity; EH, expected heterozygosity; HWE, Hardy–Weinberg *p*-values; PIC, polymorphism information content; PD, power of discrimination; PE, power of exclusion. Frequencies of new alleles are shown underlined, significant HWE *p*-values (below the Bonferroni corrected threshold of 0.0023) are shown in bold.

Table 2
"Kit specific" null alleles

Marker	Genotype in Identifiler	Genotype in Powerplex16	Null allele
D13S317	12/12	11/12	11
D13S317	10/10	10/11	11
D13S317	8/8	8/11	11
D16S539	12/14	14/14	12
D18S51	14/19	19/19	14
D18S51	14/19	14/14	19
D18S51	16/20	16/16	20

Micro-Checker software [7], using standard settings. We also used Bottleneck [8], to test for indications of heterozygote excess and heterozygote deficiency based on the Wilcoxon test results of the two phase mutation model (TPM) with a 90% SSM proportion and 10% variance. We were unable to detect significant indications for heterozygote excess, heterozygote deficiency, and presence of null alleles in this combined Bhutanese dataset.

For the markers D13S317, D16S539 and D18S51, pseudo-null alleles for either the Powerplex16 kit or the Identifiler kit have been identified during this study (i.e. individuals were typed as homozygous for a certain marker using one kit, but found to be heterozygous using the other kit). These pseudo-null alleles are listed in Table 2.

For 13 of the markers tested in this study, we compared, per locus, the allele frequency distribution from Bhutan with the allele frequency data from our Nepalese sample [9] and 13 other neighbouring populations [10–18]. For many markers, significant *p*-values (below the Bonferroni corrected threshold of 0.00004) were observed (Table 3). Since Bhutan is a relatively isolated country, mainly due to it's geography, some differences between the Bhutanese population and neighbouring populations can be expected. However, the large number of significant differences found here cannot be explained by geographical isolation only. We are currently collecting genotype data from populations in India and China, in order to perform a more to-the-point and genotype based comparison among these populations.

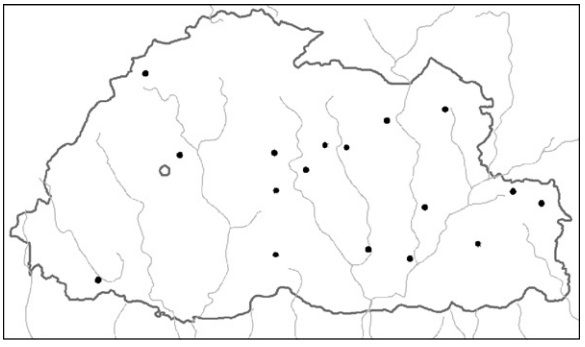

Fig. 1. Geographical centres of the 17 Bhutanese groups, defined by language (total number of individuals = 936), of which samples were included in this study. Black dots indicate the geographical centres of each group, the grey circle indicates the location of Thimphu.

Table 3
Pairwise comparison between allele frequencies from Bhutan (present study) and those from neighbouring populations (from literature)

Bhutan vs.	*n*	*n* sign.	D3S1358	VWA	D16S539	D8S1179	D21S11	D18S51	TH01	FGA	D5S818	D13S317	D7S820	CSF1PO	TPOX
Nepal [8]	953	9	0.00000	0.00360	0.00460	0.00000	0.00000	0.00000	0.02885	0.00000	0.00000	0.05360	0.00000	0.00000	0.00000
Nepali S [9]	110	10	0.00000	0.00000	0.08365	0.00000	0.00000	0.00000	0.00000	0.00000	0.26930	0.00000	0.00000	0.40855	0.00000
Luoba C [10]	93	6	0.00000	0.00000	0.00305	0.00000	0.00000	0.00000	0.94215	0.03310	0.00000	0.03170	0.00365	0.00705	0.00620
Chao Shan C [11]	144	10	0.90065	0.00000	0.00365	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00810	0.00000
Min Nan C [12]	122	5	0.04015	0.02665	0.00050	0.00000	0.00000	0.18490	0.00030	0.00000	0.00000	0.00010	0.00000	0.01950	0.00310
Baniya I [13]	90	10	0.00015	0.00000	0.15025	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00015	0.00000	0.00000
Yadav I [13]	90	9	0.00000	0.00115	0.00310	0.00000	0.00000	0.00000	0.00000	0.00000	0.00415	0.00000	0.00000	0.14435	0.00000
Bhutia S [9]	75	10	0.00925	0.00000	0.00245	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.43885	0.00000	0.00000	0.00000
Lepcha S [9]	48	8	0.00035	0.00000	0.10290	0.00000	0.00000	0.00000	0.00220	0.00000	0.00000	0.10590	0.00000	0.00000	0.00275
Jat I [14]	48	11	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00050	0.00000	0.00000	0.00200	0.00000	0.00000	0.00000
Thakur I [14]	45	13	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Lusei I [15]	92	9	0.01340	0.02445	0.00000	0.00045	0.00000	0.00000	0.00000	0.00000	0.10380	0.00000	0.00000	0.00000	0.00000
Naga I [16]	30	1	0.01780	0.13280	0.03570	0.64580	0.03475	0.64125	0.00000	0.01590	0.00085	0.00100	0.34565	0.03135	0.54110
Bangladeshi [17]	127	7	0.02875	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.09635	0.00530	0.02920	0.25455	0.00000

n, the number of individuals in the population; *n* sign., the number of markers showing significant *p*-values (below the Bonferroni corrected threshold of 0.00004) are underlined; S, from Sikkim; C, from China; I, from India.

Quality control: All laboratory procedures are accredited according to ISO17025. The FLDO also participates in the German DNA profiling group (GEDNAP) annual proficiency test and in the annual paternity testing workshop of the English Speaking Working Group of the International Society for Forensic Genetics (ISFG). Furthermore, for day-to-day quality control, laboratory internal controls, kit controls and genotypes for overlapping loci between kits are applied.

Access to data: Via electronic mail from communicating author.

Other remarks: The new alleles, detected during this population-screening study are currently being sequenced in collaboration with Dr. John Butler from the National Institute of Standards and Technology and will be added to the Short Tandem Repeat DNA Internet DataBase (STRBase [6]). In addition to new alleles, three banded allele patterns were observed for D19S433 and FES/FPS (one individual per locus). These individuals will also be further examined by the NIST and added to STRBase.

This paper was prepared according to this Journal's guidelines for publication of population data [19].

Acknowledgements

This work as part of the European Science Foundation EUROCORES Programme OMLL was supported by funds from NWO (Netherlands Organisation for Scientific Research, grant no. 231-70-001), the Arts and Humanities Research Board of Great Britain (grant no. AN/9585/APN15314) and the EC Sixth Framework Programme under Contract no. ERASCT-2003-980409. Collection of blood samples in Bhutan was carried out according to a Memorandum of Understanding between the Dzongkha Development Authority (DDA) in Bhutan and Leiden University Medical Center.

References

[1] G. van Driem, Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language (2 volumes), Brill, Leiden, The Netherlands, 2001.
[2] M. Raymond, F. Rousset, GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism, J. Hered. 86 (1995) 248–249.
[3] K. Lui, S.V. Muse, Powermarker: integrated analysis environment for genetic marker data, Bioinformatics 21 (2005) 2128–2129. http://www.statgen.ncsu.edu/powermarker/.
[4] http://www.promega.com/geneticidtools/.
[5] S. Schneider, D. Roessli, L. Excoffier, Arlequin Ver. 2.000: A Software for Population Genetic Data Analysis, Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000 http://www.anthro.unige.ch/arlequin/.
[6] C.M. Ruitberg, D.J. Reeder, J.M. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, Nucleic Acids Res. 29 (2001) 320–322. http://www.cstl.nist.gov/biotech/strbase/.
[7] C. van Oosterhout, W.F. Hutchinson, D.P.M. Wills, P. Shipley, MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data, Mol. Ecol. Notes 4 (2004) 535–538. http://www.microchecker.hull.ac.uk/.
[8] J.M. Cornuet, G. Luikart, Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data, Genetics 144 (1997) 2001–2014. http://www.ensam.inra.fr/URLB.

[9] T. Kraaijenbrink, G.L. van Driem, J.R.M.L. Opgenort, N.M. Tuladhar, P. de Knijff, Allele frequency distribution for 21 autosomal STR loci in Nepal. Forensic Sci. Int., 2006, doi:10.1016/j.forsciint.2006.02.014.
[10] V.K. Kashyap, S. Guha, R. Trivedi, Concordance study on 15 STR loci in three major population of Himalayan state Sikkim, J. Forensic Sci. 47 (2002) 1163–1167.
[11] L. Kang, S. Li, Allele frequencies of 15 STR loci of Luoba ethnic group in Tibet (Southwestern China), Forensic Sci. Int. 155 (2005) 219–221.
[12] S.P. Hu, X.J. Yu, J.W. Liu, K.L. Cai, Analysis of STR polymorphisms in the Chao Shan population in South China, Forensic Sci. Int. 147 (2005) 93–95.
[13] S.P. Hu, D.Q. Wu, X.H. Xu, J.W. Liu, B. Li, Genetic profile of 15 STR loci in the Min Nan population in Southeast China, Forensic Sci. Int. 152 (2005) 263–265.
[14] R. Ashma, V.K. Kashyap, Genetic polymorphism at 15 STR loci among three important subpopulation of Bihar, India, Forensic Sci. Int. 130 (2002) 58–62.
[15] M. Tandon, R. Trivedi, V.K. Kashyap, Genomic diversity at 15 fluorescent labeled short tandem repeat loci in few important populations of state of Uttar Pradesh, India, Forensic Sci. Int. 128 (2002) 190–195.
[16] B. Maity, S.C. Nunga, V.K. Kashyap, Genetic polymorphism revealed by 13 tetrameric and 2 pentameric STR loci in four Mongoloid tribal population, Forensic Sci. Int. 132 (2003) 216–222.
[17] V.K. Kashyap, N. Sarkar, R. Trivedi, Allele frequencies for STR loci of the Powerplex^{TM}16 multiplex system in five endogamous populations of India, Forensic Sci. Int. 126 (2002) 178–186.
[18] Y. Dobashi, A. Kido, N. Fujitani, M. Hara, R. Susukida, M. Oya, STR data for the AmpFlSTR Identifiler loci in Bangladeshi and Indonesian populations, Legal Med. 7 (2005) 222–226.
[19] P. Lincoln, A. Carracedo, Publications of population data of human polymorphism, Forensic Sci. Int. 110 (2000) 3–5.