Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International 168 (2007) 227–231
Forensic Science International
www.elsevier.com/locate/forsciint

Announcement of Population Data

Allele frequency distribution for 21 autosomal STR loci in Nepal

T. Kraaijenbrink [a], G.L. van Driem [b], J.R.M.L. Opgenort [b], N.M. Tuladhar [c], P. de Knijff [a,*]

[a] MGC Department of Human and Clinical Genetics, Leiden University Medical Centre, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands
[b] Himalayan Languages Project, FBJ Kuiperinstituut, Leiden University, P.N. van Eyckhof 3, 2311 BV Leiden, The Netherlands
[c] Centre for Nepal and Asian Studies (CNAS), Tribhuvan University (TU), Kirtipur, Nepal

Received 21 November 2005; received in revised form 3 February 2006; accepted 5 February 2006
Available online 15 March 2006

Abstract

The allele frequency distributions of 21 autosomal loci contained in the AmpF*l*STR® Identifiler™, the Powerplex®16 and the FFFL® multiplex PCR kits, was studied in 953 unrelated individuals from Nepal. Several new alleles (i.e. not yet reported in the NIST Short Tandem Repeat DNA Internet DataBase [http://www.cstl.nist.gov/biotech/strbase/]) have been detected in the process.

Keywords: Nepal; Identifiler; Powerplex16; FFFL

Population information: The Himalayan mountain range, running from Pakistan in the west to north-eastern India (Assam) in the east, is not only the highest land barrier on the face of this planet, it also forms a barrier between two distinct language families; Tibeto-Burman, which is predominantly spoken north-east of the Himalayas, and Indo-European, which is spoken south of the Himalayas [2]. In a number of countries such as Nepal and Bhutan, with a geographical position just south of the highest Himalayan mountain peaks, people predominantly speak languages belonging to the Tibeto-Burman language family. This suggests 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 Nepal and Bhutan. To our knowledge, there are only a handful of published genetic studies among Nepalese populations [3–7]. None of these describe autosomal STR frequency data. In order to be able to reconstruct possible migration routes explaining the earliest settling of modern humans in Nepal, we collected blood samples from 786 unrelated males and 185 unrelated females from 12 major Nepalese groups, defined by caste, religion, or language (see Fig. 1). Here, we describe the first results of our study in the form of the allele frequency distributions and summary statistics of 21 different autosomal STR loci from the combined group of 953 Nepalese individuals.

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, after which 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 [8], exact Hardy–Weinberg *p*-values and PIC values were calculated using PowerMarker [9]. The Excel PowerStats spreadsheet [10] was used to calculate power of discrimination and power of exclusion. We used the exact test of population differentiation included in Arlequin Version 2000 to compare differences in allele frequencies among different populations. Because we were unable to obtain raw genotype scores from populations not typed by us, we had to base these test on allele frequency data.

Results: Allele frequencies and summary statistics are presented in Table 1. We identified a number of new alleles. These are shown underlined in Table 1. Using the PowerMarker software, markers Penta D, TPOX and F13A01 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

* Corresponding author. Tel.: +31 71 527 4318; fax: +31 71 527 4517.
E-mail address: knijff@lumc.nl (P. de Knijff).

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2006.02.014

Table 1
Allele frequencies for 21 STR loci in 953 unrelated Nepalese samples

Alleles	D3S1358	VWA	D16S539	D8S1179	D21S11	D18S51	TH01	FGA	D2S1338	D19S433	PENTA E	D5S818	D13S317	D7S820	CSF1PO	PENTA D	TPOX	F13A01	F13B	FES/FPS	LPL
3.2																		0.2193			
3.3																		0.0005			
4																		0.1469			
5											0.0519					0.0005		0.1936			
6							0.0871									0.0010		0.3841	0.0199		
7							0.2183				0.0273	0.0110	0.0021	0.0136	0.0005	0.0147	0.0031	0.0294	0.0010	0.0005	
8			0.0236	0.0042			0.0992				0.0089	0.0021	0.1605	0.2267	0.0016	0.0530	0.4517	0.0010	0.1469	0.0037	0.0005
8.1														0.0005							
8.2														0.0010							
9			0.3027	0.0026			0.4848				0.0079	0.1018	0.1626	0.0787	0.0645	0.2256	0.1537		0.2225	0.0052	0.0084
9.3							0.1013														
10			0.0992	0.0923		0.0021	0.0094				0.0241	0.1878	0.1647	0.1459	0.1705	0.1621	0.0357		0.6049	0.0756	0.6201
10.1														0.0037							
11			0.2733	0.0509		0.0079				0.0026	0.1417	0.3263	0.2497	0.2545	0.2692	0.2461	0.3316	0.0005	0.0037	0.4192	0.1144
11.2										0.0005											
11.3										0.0005											
12		0.0005	0.2104	0.1296		0.0456				0.0414	0.1128	0.2088	0.1962	0.2408	0.4029	0.1390	0.0152		0.0010	0.2749	0.2078
12.2										0.0063											
13		0.0005	0.0797	0.2051		0.1952				0.2954	0.0467	0.1569	0.0514	0.0320	0.0740	0.1144	0.0079	0.0016		0.1988	0.0456
13.1										0.0005											
13.2										0.0456											
14	0.0362	0.1485	0.0105	0.2282		0.2057				0.2812	0.0923	0.0052	0.0126	0.0026	0.0152	0.0304	0.0010	0.0037		0.0215	0.0021
14.2										0.0860											
15	0.3237	0.0289	0.0005	0.1925		0.1469				0.0845	0.1112				0.0016	0.0121		0.0115		0.0005	0.0010
15.2										0.1039	0.0005										
16	0.3478	0.2650		0.0766		0.1228			0.0042	0.0262	0.1233					0.0010		0.0068			
16.2	0.0005									0.0220											
17	0.1962	0.2786		0.0178		0.0493		0.0005	0.0493	0.0021	0.0703							0.0010			
17.2										0.0010											
18	0.0855	0.1695				0.0446		0.0399	0.1180		0.0577										
19	0.0089	0.0892				0.0834		0.0745	0.1899		0.0525										
20	0.0010	0.0189				0.0378		0.0645	0.1186		0.0315										
20.2								0.0021													
21						0.0215		0.0693	0.0409		0.0178										
21.2								0.0089													
22		0.0005				0.0110		0.1118	0.0519		0.0094										
22.2								0.0073													
23						0.0147		0.1962	0.1784		0.0063										
23.2								0.0068													
24						0.0068		0.2046	0.1474		0.0042										
24.2								0.0110													
25						0.0031		0.1175	0.0866		0.0016										
25.2								0.0068													
26					0.0005	0.0010		0.0488	0.0131												
26.2								0.0010													
27					0.0068			0.0184	0.0005												
27.2					0.0005																
28					0.0771	0.0005		0.0084	0.0010												

28.2				0.0283																
29				0.2560																
29.2				0.0031		0.0016														
30				0.2560																
30.2				0.0157																
31				0.0923																
31.2				0.0714																
32				0.0152																
32.2				0.1201																
33				0.0010																
33.1				0.0010																
33.2				0.0483																
34				0.0005																
34.1				0.0010																
34.2				0.0042																
35.2				0.0005																
OH	0.7188	0.7692	0.7639	0.8258	0.8216	0.8311	0.6516	0.8625	0.8510	0.7639	0.8762	0.7524	0.8017	0.7754	0.6915	0.7891	0.6128	0.6915	0.5467	0.6716	0.5310
EH	0.7274	0.7927	0.7729	0.8351	0.8316	0.8676	0.6899	0.8745	0.8683	0.8037	0.9128	0.7799	0.8174	0.7976	0.7266	0.8262	0.6611	0.7447	0.5629	0.7033	0.5573
HWE	0.4541	0.0711	0.6676	0.0754	0.4981	0.1399	0.0054	0.1967	0.0183	0.1167	0.0112	0.3809	0.1387	0.0104	0.0179	**0.0016**	**0.0029**	**0.0010**	0.5812	0.4091	0.0073
PIC	0.6796	0.7615	0.7368	0.8135	0.8117	0.8535	0.6510	0.8619	0.8537	0.7787	0.9057	0.7461	0.7909	0.7667	0.6838	0.8029	0.6000	0.7048	0.5081	0.6529	0.5107
PD	0.8757	0.9251	0.9142	0.9517	0.9506	0.9683	0.8626	0.9706	0.9670	0.9391	0.9847	0.9196	0.9402	0.9295	0.8838	0.9476	0.8294	0.8964	0.7541	0.8657	0.7559
PE	0.4579	0.5431	0.5339	0.6478	0.6398	0.6579	0.3575	0.7197	0.6968	0.5339	0.7471	0.5138	0.6022	0.5543	0.4152	0.5789	0.3065	0.4152	0.2317	0.3857	0.2161

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
Amelogenin	–	XY	X and Y
D13S317	14/14	12/14	12
D13S317	11/11	11/12	12
D16S539	12/14	14/14	12
D16S539	8/12	8/8	12
D16S539	10/12	10/10	12
D18S51	18/19	18/18	19
D18S51	16/19	16/16	19
D18S51	17/19	17/17	19
D18S51	14/19	14/14	19
D18S51	14/19	14/14	19
D18S51	16/19	16/16	19
D18S51	19/21	21/21	19
D18S51	14/19	14/14	19
D21S11	33.2/34.2	34.2/34.2	33.2
D7S820	8/8.2	8/8	8.2
D7S820	8/8.2	8/8	8.2
FGA	23/24.2	23/23	24.2
TPOX	8/10	8/8	10

significant heterozygote deficiency for markers Penta D and F13A01. Subsequently, in order to test for the presence of undetected null alleles as a cause for heterozygote deficiency, the data were analysed using the Micro-Checker software [11], using standard settings. We also used Bottleneck [12] 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 Nepalese dataset.

For the markers Amelogenin, D13S317, D16S539, D18S51, D21S11, D7S820, FGA and TPOX, pseudo-null alleles for the Powerplex16 kit and the Identifiler kit were discovered (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.

The allele frequency data obtained in this study were compared to allele frequency data from a number of neighbouring populations [13–21]. Many significant differences were found between our Nepalese sample and neighbouring populations (Table 3). Although some significant differences between Nepal and its neighbours are to be expected, the relative high number of significant differences is surprising, and cannot easily be explained by geographical distance or linguistic affiliations. In this respect, it is important to realise that comparing allele frequency data of individual loci among populations is a rather simplistic approach, rarely revealing the relevant genetic differences among populations. A meaningful comparison with other populations should preferably be based on genotypic data, which allows the use of better analytical tools. In order to be able to do so, we are currently collecting genotype data from populations in India and China, in collaboration with groups in those countries.

Quality control: All laboratory procedures are accredited according to ISO17025. The FLDO also participates in the

Table 3
Pairwise comparison of the allele frequency data from the current study to allele frequency data from neighbouring populations, obtained from literature

Nepal vs. . . .	*n*	*n* sign.	D3S1358	VWA	D16S539	D8S1179	D21S11	D18S51	TH01	FGA	D5S818	D13S317	D7S820	CSF1PO	TPOX
Nepali **S** [13]	110	7	0.00000	0.00000	0.00000	0.00000	0.02160	0.18670	0.00010	0.00000	0.16420	0.00000	0.00810	0.88425	0.00000
Luoba **C** [14]	93	5	0.00000	0.00010	0.00020	0.00000	0.00000	0.00000	0.46990	0.01725	0.00000	0.00005	0.06045	0.03935	0.04955
Chao Shan **C** [15]	144	6	0.07590	0.00000	0.08525	0.03275	0.00000	0.00710	0.00000	0.00000	0.00440	0.00000	0.00000	0.00330	0.01445
Min Nan **C** [16]	122	5	0.42670	0.00050	0.07140	0.00000	0.02205	0.00000	0.00030	0.00000	0.01320	0.00000	0.00000	0.02720	0.05435
Baniya **I** [17]	90	11	0.00000	0.00000	0.01035	0.00035	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Yadav **I** [17]	90	10	0.00000	0.00650	0.00250	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06830	0.00000
Bhutia **S** [13]	75	9	0.02250	0.00000	0.00485	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.05935	0.00000	0.00055	0.00000
Lepcha **S** [13]	48	6	0.05380	0.00000	0.00145	0.00000	0.00000	0.00155	0.00325	0.00000	0.00000	0.17890	0.00015	0.00000	0.00535
Jat **I** [18]	48	7	0.00000	0.00000	0.00000	0.03810	0.00460	0.00000	0.00070	0.00000	0.00000	0.00325	0.00015	0.00165	0.00000
Thakur **I** [18]	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** [19]	92	8	0.02440	0.31790	0.00000	0.00000	0.00000	0.00000	0.00040	0.00000	0.00165	0.00000	0.03475	0.00000	0.00000
Naga **I** [20]	30	1	0.02560	0.22385	0.02825	0.89015	0.00295	0.64350	0.00000	0.00010	0.01395	0.00035	0.15390	0.06665	0.81900
Bangladeshi [21]	127	4	0.92170	0.00030	0.00000	0.00060	0.00000	0.00000	0.00000	0.00855	0.00950	0.04140	0.01990	0.24745	0.00140

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.

Fig. 1. Geographical centres of the 12 major Nepalese groups (total number of individuals = 953), predominantly defined by language differentiation, of which samples were included in this study. Black dots indicate the geographical centres and the grey circle indicates the location of Kathmandu.

German DNA profiling group (GEDNAP) and the International Society for Forensic Genetics (ISFG) annual proficiency tests. Furthermore, for day-to-day quality control, laboratory internal controls, kit controls and genotypes for overlapping loci between kits are applied.

Access to data: Through electronic mail from communicating author.

Other remarks: The new alleles and the pseudo-null 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) [1].

This paper was prepared according to this Journal's guidelines for publication of population data [22].

Acknowledgements

This work, as part of the European Science Foundation EUROCORES Programme OMLL, was also 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. ERAS-CT-2003-980409.

References

[1] http://www.cstl.nist.gov/biotech/strbase/.
[2] 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 vols.), Brill, Leiden, The Netherlands, 2001.
[3] H. Pang, Y. Koda, M. Soejima, N. Fujitani, T. Ogaki, A. Saito, T. Kawasaki, H. Kimura, Polymorphism of the human ABO-Secretor locus (FUT2) in four populations in Asia: indication of distinct Asian subpopulations, Ann. Hum. Genet. 65 (2001) 429–437.
[4] S. Umemura, T. Kawasaki, T. Ishigami, T. Fujita, K. Hibi, M. Kawasaki, K. Itoh, Y. Yoshimizu, T. Ogaki, G.P. Acharya, M. Ishii, Angiotensin-converting enzyme gene polymorphism in Nepal, J. Hum. Hypertens. 12 (1998) 527–531.
[5] G. Passarino, O. Semino, G. Pepe, S.L. Shrestha, G. Modiano, A.S. Santachiara Benerecetti, MtDNA polymorphisms among Tharus of eastern Terai (Nepal), Gene Geogr. 6 (1992) 139–147.

[6] O. Semino, A. Torroni, R. Scozzari, A. Brega, A.S. Santachiara Benerecetti, Mitochondrial DNA polymorphisms among Hindus: a comparison with the Tharus of Nepal, Ann. Hum. Genet. 55 (1991) 123–136.
[7] C.R. Bangham, S.E. Howarth, Genetic polymorphisms in isolated Sherpa populations of Nepal, Am. J. Phys. Anthropol. 53 (1980) 369–373.
[8] M. Raymond, F. Rousset, GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism, J. Hered. 86 (1995) 248–249.
[9] http://statgen.ncsu.edu/powermarker/.
[10] http://ww.promega.com/geneticidtools/.
[11] http://www.microchecker.hull.ac.uk/.
[12] http://www.montpellier.inra.fr/URLB/.
[13] 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.
[14] 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.
[15] 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.
[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] V.K. Kashyap, N. Sarkar, R. Trivedi, Allele frequencies for STR loci of the Powerplex™16 multiplex system in five endogamous populations of India, Forensic Sci. Int. 126 (2002) 178–186.
[21] Y. Dobashi, A. Kido, N. Fujitani, M. Hara, R. Susukida, M. Oya, STR data for the AmpF*l*STR Identifiler loci in Bangladeshi and Indonesian populations, Legal Med. 7 (2005) 222–226.
[22] P. Lincoln, A. Carracedo, Publications of population data of human polymorphism, Forensic Sci. Int. 110 (2000) 3–5.