A genetic database for DNA-based forensic analysis in Brunei Darussalam

Poh Yee Cheong [1], Paul V.O. Liew [1], Hamydon Ibrahim [1], Bruce Budowle [2]

[1] *Forensic Biology/DNA Laboratory, Department of Scientific Services, Ministry of Health Complex, Commonwealth Drive, Jalan Menteri Besar, Bandar Seri Begawan, Brunei Darussalam BB3910*
[2] *Federal Bureau of Investigation, Laboratory Division, Quantico, VA 22135, USA*

Abstract

Allele frequency data for the 9 short tandem repeat (STR) loci, D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820 were determined in the Malay and the Chinese populations in Brunei Darussalam. The combined power of discrimination using these loci was determined to be >0.99999999 for both the Malay and Chinese populations. The combined power of exclusion was determined to be >0.9998 for the Malays and > 0.9999 for the Chinese. The Brunei Darussalam Malay and Chinese database for the 9 STR loci is therefore useful for forensic human identification and parentage testing.

Keywords: Brunei Darussalam population data, Chinese, forensic DNA typing, Malay, polymerase chain reaction, short tandem repeats

Introduction

The Department of Scientific Services, Ministry of Health is in the process of establishing forensic DNA testing for the first time in Brunei Darussalam. If the DNA profile determined from samples at a crime scene matches that of the suspect, they may have come from the same person, or they might be a coincidental match between two persons in the population who happen to share the profile. The purpose of a DNA population frequency database is to permit statistical interpretation of DNA profiling results to assess the probability of a coincidental match. This information determines the strength of the DNA-based forensic evidence in a court of law.

Short tandem repeats (STR) markers are polymorphic DNA loci that contain a repeated nucleotide sequence with 2 – 6 bp in length and they are the most informative PCR-based genetic markers for attempting to individualise biological material. Different allelic forms at each loci are composed of different numbers of repeats. In 1998, the Federal Bureau of Investigation of the United States (FBI) selected 13 such STR loci as the basis for a national database to identify criminals termed the Combined DNA Index System (CODIS). A commercial kit that used multiplex PCR of nine of these STR markers (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820) in one reaction, together with size analysis of the PCR products by capillary electrophoresis for determining the different allelic forms, was selected for establishing a Bruneian DNA-type frequency database.

Correspondence:
Poh-Yee Cheong, M.Sc., Acting Director, Department of Scientific Services, Ministry of Health Complex, Commonwealth Drive, Jalan Menteri Besar, Bandar Seri Begawan, Brunei Darussalam BB3910
Tel: +673 2381899
Fax: +673 2381946
Email address: pohyee6@hotmail.com

Materials and Methods

Blood samples from unrelated and healthy individuals in Brunei Darussalam were collected from blood bank donors and volunteers in the Brunei-Muara and Tutong Districts. The sample population consisted of 167 Malays and 125 Chinese. The blood samples were collected as anonymous samples. Clearance of the study was obtained from the director of the department since there was no research ethical review board available at the time of study. Individuals who provided blood samples were asked to respond to a questionnaire about detailed ethnic information indicating racial pedigree over a period of 4 generations. The criterion to include individual DNA data into the nominated racial group was the selection only of Bruneian individuals who indicated greater than three-quarter heritage of that racial group.

DNA was extracted from blood samples stained on FTA Cards (Whatman Bioscience, MA, USA) according to the manufacturer's protocol (Applied Biosystems AmpFlSTR® Profiler Plus™ User's Manual). Some of the DNA was also obtained from standardized 1mm punch out blood samples stained on Ultrastain Cards (Whatman Bioscience, MA, USA) using the Chelex (Bio-Rad, CA, USA) extraction protocol [1]. The extracted DNA was amplified directly using the AmpFlSTR® Profiler Plus™ Amplification Kits (Applied Biosystems, Foster City, USA) on a GeneAmp 9700 PCR system (Applied Biosystems, Foster City, USA) as described by the manufacturer's instruction. The analysis of the amplified PCR product was performed on the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, USA). The identification of the alleles was analyzed using the GeneScan (Version 3.1.2) and Genotyper (Version 2.5) software (Applied Biosystems, Foster City USA).

Allele designations were determined by comparison of the sample fragments with those of the allelic ladders. The frequency of each allele for each locus was calculated from the numbers of each genotype in the sample set. In other words, allele frequency determination was made by the gene count method [2], which does not require any assumption regarding population substructure. Unbiased estimates of expected heterozygosity were computed [3]. Possible divergence from Hardy-Weinberg Equilibrium was tested by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies [3-5] and the exact test [6], based on 2000 shuffling experiments. An inter-class correlation criterion for two-locus associations was used for detecting disequilibrium between the STR loci [7]. The power of discrimination and power of exclusion were performed as described by Fisher [8]. The computer programme to perform these tests was developed by R. Chakraborty (Department of Environmental Health, University of Cincinnati, Ohio, USA, http://cgi.uc.edu/).

Results and Discussion

The frequency distributions of the observed allele for the 9 short tandem repeat (STR) loci and other forensic related statistics for the Malay and Chinese populations in Brunei Darussalam are shown in Tables 1 and 2 respectively.

Table 1. Observed STR allele frequencies for the 9 STR loci for Brunei Darussalam Malay population (N=167)

Allele	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
7	-	-	-	-	-	-	0.0060	0.0030	0.0060
8	-	-	-	-	-	-	0.0090	0.2515	0.2066
9	-	-	-	0.0030	-	0.0030	0.0509	0.0958	0.0509
10	-	-	-	0.0659	-	0.0030	0.3024	0.1467	0.2275
11	-	-	-	0.1617	-	0.0060	0.2275	0.3024	0.3533
12	-	-	-	0.0808	-	0.0659	0.2844	0.1796	0.1347
13	-	-	-	0.2395	-	0.0689	0.1078	0.0150	0.0150
14	0.0180	0.1677	-	0.1856	-	0.1976	0.0090	0.0060	0.0060
15	0.2665	0.1018	-	0.1677	-	0.3174	0.0030	-	-
16	0.3323	0.1317	-	0.0928	-	0.1258	-	-	-
17	0.3054	0.2335	-	0.0030	-	0.0659	-	-	-
<18	-	-	-	-	-	-	-	-	-
18	0.0719	0.2784	0.0180	-	-	0.0449	-	-	-
19	0.0060	0.0778	0.0629	-	-	0.0539	-	-	-

19.2	-	-	0.0030	-	-	-	-	-	-
20	-	0.0090	0.0299	-	-	0.0150	-	-	-
20.2	-	-	0.0030	-	-	-	-	-	-
21	-	-	0.1946	-	-	0.0180	-	-	-
21.2	-	-	0.0030	-	-	-	-	-	-
22	-	-	0.1767	-	-	-	-	-	-
22.2	-	-	0.0180	-	-	-	-	-	-
23	-	-	0.1737	-	-	0.0030	-	-	-
23.2	-	-	0.0060	-	-	-	-	-	-
24	-	-	0.1228	-	-	0.0060	-	-	-
24.2	-	-	0.0030	-	-	-	-	-	-
25	-	-	0.0838	-	0.0030	0.0060	-	-	-
25.2	-	-	0.0090	-	-	-	-	-	-
26	-	-	0.0569	-	-	-	-	-	-
26.2	-	-	0.0030	-	-	-	-	-	-
27	-	-	0.0240	-	-	-	-	-	-
28	-	-	0.0090	-	0.0450	-	-	-	-
28.2	-	-	-	-	0.0030	-	-	-	-
29	-	-	-	-	0.2275	-	-	-	-
30	-	-	-	-	0.2305	-	-	-	-
30.2	-	-	-	-	0.0270	-	-	-	-
31	-	-	-	-	0.1856	-	-	-	-
31.2	-	-	-	-	0.0569	-	-	-	-
32	-	-	-	-	0.0419	-	-	-	-
32.2	-	-	-	-	0.1198	-	-	-	-
33	-	-	-	-	0.0030	-	-	-	-
33.2	-	-	-	-	0.0449	-	-	-	-
34	-	-	-	-	0.0030	-	-	-	-
34.2	-	-	-	-	0.0060	-	-	-	-
35.2	-	-	-	-	0.0030	-	-	-	-
H (obs)	0.234	0.192	0.156	0.210	0.144	0.144	0.198	0.257	0.257
H (exp)	0.278	0.192	0.128	0.163	0.161	0.172	0.236	0.216	0.238
P	0.746	0.360	0.734	0.489	0.479	0.963	0.701	0.435	0.479
PD	0.856	0.929	0.968	0.950	0.950	0.949	0.894	0.919	0.905
PE	0.466	0.617	0.739	0.668	0.677	0.668	0.539	0.574	0.541

H (obs): observed homozygosity ;

H (exp): expected homozygosity ;

P: P-value of the exact tests for Hardy-Weinberg equilibrium ;

PD: power of discrimination ;

PE: probability of excluding paternity.

Table 2. Observed STR allele frequencies for the 9 STR loci for Brunei Darussalam Chinese population (N=125)

Allele	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
7	-	-	-	-	-	-	0.0320	-	0.0080
8	-	-	-	-	-	-	0.0160	0.2800	0.1320
9	-	-	-	-	-	-	0.0680	0.1280	0.0800
10	-	-	-	0.1320	-	0.0040	0.2040	0.1520	0.1400
11	-	-	-	0.0960	-	0.0040	0.3120	0.2640	0.3800
12	-	-	-	0.1640	-	0.0520	0.2120	0.1120	0.2160
13	-	-	-	0.2280	-	0.1720	0.1560	0.0520	0.0360
14	0.0320	0.2680	-	0.1720	-	0.1800	-	0.0120	0.0080
15	0.3560	0.0320	-	0.1480	-	0.1800	-	-	-
16	0.2800	0.1240	-	0.0440	-	0.1480	-	-	-
17	0.2640	0.2800	-	0.0160	-	0.0800	-	-	-
<18	-	-	0.0040	-	-	-	-	-	-
18	0.0640	0.1880	0.0560	-	-	0.0480	-	-	-
19	0.0040	0.0880	0.0400	-	-	0.0640	-	-	-
19.2	-	-	-	-	-	-	-	-	-
20	-	0.0200	0.0840	-	-	0.0200	-	-	-
20.2	-	-	-	-	-	-	-	-	-
21	-	-	0.1240	-	-	0.0160	-	-	-
21.2	-	-	0.0040	-	-	-	-	-	-
22	-	-	0.1560	-	-	0.0280	-	-	-
22.2	-	-	0.0040	-	-	-	-	-	-
23	-	-	0.2160	-	-	0.0040	-	-	-
23.2	-	-	-	-	-	-	-	-	-
24	-	-	0.1400	-	-	-	-	-	-
24.2	-	-	0.0080	-	-	-	-	-	-
25	-	-	0.0920	-	-	-	-	-	-
25.2	-	-	-	-	-	-	-	-	-
26	-	-	0.0520	-	-	-	-	-	-
26.2	-	-	0.0160	-	-	-	-	-	-
27	-	-	0.0040	-	-	-	-	-	-
28	-	-	-	-	0.0680	-	-	-	-
28.2	-	-	-	-	0.0040	-	-	-	-
29	-	-	-	-	0.2360	-	-	-	-
30	-	-	-	-	0.2640	-	-	-	-
30.2	-	-	-	-	-	-	-	-	-
31	-	-	-	-	0.0960	-	-	-	-
31.2	-	-	-	-	0.0760	-	-	-	-
32	-	-	-	-	0.0160	-	-	-	-
32.2	-	-	-	-	0.1560	-	-	-	-
33	-	-	-	-	0.0080	-	-	-	-
33.2	-	-	-	-	0.0560	-	-	-	-
34	-	-	-	-	-	-	-	-	-

34.2	-	-	-	-	0.0200	-	-	-	-
35.2	-	-	-	-	-	-	-	-	-
H (obs)	0.232	0.216	0.112	0.216	0.128	0.168	0.200	0.200	0.240
H (exp)	0.277	0.207	0.126	0.156	0.170	0.130	0.211	0.200	0.233
P	0.282	0.961	0.508	0.510	0.362	0.598	0.706	0.764	0.181
PD	0.860	0.920	0.962	0.952	0.931	0.964	0.918	0.927	0.908
PE	0.469	0.589	0.741	0.679	0.661	0.733	0.582	0.603	0.559

H (obs): observed homozygosity ;
H (exp): expected homozygosity ;
P: P-value of the exact tests for Hardy-Weinberg equilibrium ;
PD: power of discrimination ;
PE: probability of excluding paternity.

None of the 9 loci were found to deviate significantly from Hardy-Weinberg expectations. The overall results of the statistical analysis suggest that the allele frequency data in the Malay and Chinese populations are reasonable and valid for estimating the rarity of a multiple loci profile.

The inter-class correlation test analysis [7, 8] performed to detect any correlations between alleles used 9 loci for pair-wise comparisons in both the Malay and the Chinese population database. For each sample population, there were a total of 36 pair-wise comparisons performed. For the Malay database, the number of significant departures was one observation (2.8%), which is D18S51/D5S818 ($p = 0.033$). The number of observed departures for the Malay dataset is no more than would be expected by chance. For the Chinese database, there were four significant departures (11.1%), namely vWA/813S317 ($p = 0.48$), vWA/D7S820 ($p = 0.019$), FGA/D8S1179 ($p = 0.027$) and D8S1179/D13S317 ($p = 0.020$). The number observed in the Chinese dataset is higher than expected. However, after Bonferroni correction (9), these departures were no longer significant. The combined power of discrimination was > 0.99999999 for both the Malay and Chinese populations. The combined power of exclusion was > 0.9998 for the Malay population and > 0.9999 for the Chinese population.

To further assess the reliability of the Malay and Chinese population data in Brunei Darussalam, the allele frequencies were compared using an RXC test [10, 11] with other similar populations [12, 13]. The commonly occurring alleles are the same in the respective population comparisons. There were no significant differences.

In conclusion the results suggest that the database based on the 9 STR loci established for the Malay and Chinese populations in Brunei Darussalam has the requisite discriminative ability for use in forensic human identification and for establishing parentage. The complete data set is available to any interested researcher upon request from the corresponding author.

Acknowledgements

The authors thank the mobile staff of the Blood Bank, RIPAS Hospital, and the staff of the Department of Scientific Services in Bandar Seri Begawan, Brunei Darussalam for their assistance in the collection of the blood samples.

References

1. Walsh PS, Meztger DA, Higuchi R 1991. Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10: 506-513.

2. Li CC 1976. First Course in Population Genetics, Boxwood Press, Pacific Grove, CA, First Edition.

3. Chakraborty R, Smouse PE, Neel JV 1988. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. Amer J Hum Genet 43: 709-725.

4. Nei M, Roychoudhury AK 1974. Sampling variances of heterozygosity and genetic distance. Genetics 76: 379-390.

5. Nei M 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

6. Guo SW, Thompson EA 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48: 361-372.

7. Karlin S, Cameron EC, Williams PT 1981. Sibling and parent-offspring correlation estimation with variable family size. Proc Natl Acad Sci USA 78: 2664-2668.

8. Fisher, RA 1951. Standard calculations for evaluating a blood group system. Heredity 5: 95-102.

9. Weir BS 1990. "Multiple tests," in Genetic Data Analysis, Sinauer Associates, Inc., Sunderland, Massachusetts pp 109-110.

10. Lewontin RC, Felsenstein J 1965. The robustness of homogeneity tests in 2 X N Tables. Biometrics 21: 19–33.

11. Roff DA, Bentzen P 1989. The statistical analysis of mitochondrial DNA polymorphisms: and the problem of small samples. Mol Biol Evol 6: 539–545.

12. Ang HC, Sornarajah R, Lim SES, Syn CKC, Tan-Siew WF, Chow ST, et al. 2005. STR data for the 13 CODIS loci in Singapore Malays. Forensic Sc Int. 148: 243-245.

13. Syn CKC, Chuah SY, Ang HC, Lim SES, Tan-Siew WF, Chow ST, et al. 2005. Genetic data for the 13 CODIS STR loci in Singapore Chinese. Forensic Sc. Int 152: 285-288.