Population Genetics of Identifiler System in Malaysia

Yasutaka Nakamura1, Michinaga Samejima2, Kiyoshi Minaguchi3 and Phrabhakaran Nambiar4

1) Department of Forensic Odontology and Forensic Anthropology, Tokyo Dental College, 2-9-18 Misaki-cho, Chiyoda-ku, Tokyo 101-0061, Japan
2) Private Practice, Hidaka-gun, Hokkaido, Japan
3) Department of Forensic Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara-shi, Kanagawa 259-1193, Japan
4) Department of General Dental Practice and Oral & Maxillofacial Imaging, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia

Received 16 February, 2016/Accepted for publication 6 May, 2016

Abstract

Short tandem repeat (STR) polymorphisms were investigated in 341 unrelated Malay individuals (218 males and 123 females) living in or around Kuala Lumpur by using a forensic analysts kit. The following STRs were targeted: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA. The purpose of this study was to elucidate population genetics in Malaysia and calculate statistical parameters for forensic and anthropological research. Data on these STRs in the target population were obtained and subjected to statistical analysis. Accordance with the Hardy-Weinberg equilibrium was proven for all the loci targeted. The combined power of discrimination was greater than 0.9999999999, indicating that this multiplex system is an excellent tool for forensic casework. The allele frequency in the data were weighed against that in four other local populations (Chinese, Iranian, Belgian, and African). The average coefficient of correlation was strongest in the order of Africa (0.092522), Belgium (0.264822), Iran (0.404363), and China (0.706661). These results are consistent with what is known about the anthropological history of and prehistoric human migration in the Malay region. We believe that these data offer a valuable anthropological resource, being applicable to the statistical evaluation of DNA evidence in human identification, as well as the determination of ethnicity in healthy populations.

Key words: DNA identification — STR — Malaysia
Introduction

Human DNA has been the focus of much research in the fields of population genetics and forensics in recent years, with analysis of short tandem repeats (STRs), in particular, being a common tool in such work worldwide. The United States Federal Bureau of Investigation has approved the Applied Biosystems Identifiler® Plus forensic kit in generating DNA profiles for inclusion in the National DNA Index System Database. This kit is now used widely throughout the world, and is also the one chosen by the Japanese police.

Data collection for such studies is conducted based on many populations\textsuperscript{1–3,5,7}, and it is essential that such DNA analysis be as accurate as possible, which means allele frequency must be established in making an identification. From an anthropological point of view, Southeast Asia is an important region. The Asian race spread out from Southeast Asia into East Asia (Japan, Korea, and China). Therefore, population data on local Southeast Asian populations is important in applying DNA analysis to criminal cases where identification may involve foreign nationals.

The purpose of this study was to elucidate population genetics in Malaysia by using the Identifiler system and to calculate statistical parameters for forensic use.

Materials and Methods

1. Samples
Genomic DNA was extracted from tooth samples obtained from 341 unrelated Malay individuals (218 males and 123 females) living in or around Kuala Lumpur. The Malaysian population comprises 65% of individuals of Malay Archipelago, 24% of Chinese, and 8% of Indian origin. This study used data exclusively of Malay Archipelago origin.

Appropriate consent was obtained from the participants. A family history was also taken to ensure that the parents were of Malay origin. No tribal population samples were included in the study. This study was approved by the Ethics Committee of Tokyo Dental College (approval no. 202 and 204) and met the conditions for cooperative study at the University of Malaya.

2. DNA extraction
Isolation of genomic DNA from tooth and cotton samples was performed as described previously\textsuperscript{8}. Permanent teeth extracted during dental treatment were used to obtain each sample. Each tooth was crushed to facilitate extraction of DNA.

3. PCR
The Investigator Identifiler® Plus Amplification kit (Applied Biosystems) was used for PCR amplification and typing of autosomal STRs at the following loci: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA. Table 1 shows the positional information on each STR. The primer information of this kit is not disclosed.

The ABI PRIZM 3130 Genetic Analyzer (Applied Biosystems) was used to perform

| Locus designation | Chromosome location |
|-------------------|---------------------|
| D8S1179           | 8                   |
| D21S11            | 21q11.2- q21        |
| D7S820            | 7q11.21- 22         |
| CSF1PO            | 5q33.3- 34          |
| D3S1358           | 3p                  |
| TH01              | 11p15.5             |
| D13S317           | 12q22-31            |
| D16S539           | 16q24-qter          |
| D2S1338           | 2q35-37.1           |
| D19S433           | 19q35-37.1          |
| vWA               | 12p12-pter          |
| TPOX              | 2q23-2per           |
| D18S51            | 18q21.3             |
| D5S818            | 5p21-31             |
| FGA               | 4q28                |
electrophoresis under the conditions described in the manufacturer’s recommendations. Fragment sizes were automatically determined using the GeneScan Analysis software 3.1 (Applied Biosystems) and the results analyzed using Genotyper version 2.5 (Applied Biosystems).

4. Statistical analysis

Linkage disequilibrium and the Hardy-Weinberg equilibrium (H-W, p-value) were determined with the exact test using the GENEPOP software version 4.2 (http://genepop.curtin.edu.au). The power of discrimination (PD), power of exclusion (PE), typical paternity index (TPI), polymorphic information contents (PIC), and observed heterozygosity (HT) were also calculated with the PowerStatsV12 software (http://www.promega.com).

Results

Tables 2A, 2B, 2C, and 2D show the data obtained on the allele frequencies of the 15 STRs loci targeted. The statistical parameters are shown in Table 3.

Possible deviations from the H-W equilibrium were determined with the exact test at a
5% level of significance. The p-value ranged from 0.0668 (TH01) to 0.9235 (CSF1PO). Based on the results of the exact test, no deviation from the p-value was observed at any of the 15 loci. The PD ranged from 0.793 (TPOX) to 0.971 (FGA); the PE from 0.337 (TPOX) to 0.754 (FGA); the TPI from 1.38 (TPOX) to 4.16 (FGA); the PIC from 0.57 (TPOX) to 0.87 (FGA); and the HT from 0.636 (TPOX) to 0.880 (FGA). In addition, the total number of alleles checked in all the STRs was 682.

**Discussion**

The statistical parameters revealed a high rate of polymorphisms at all loci targeted, and a particularly high rate at 12 of these (PD >0.9). The most informative however, appear to be the following loci: D8S1179, D21S11, D18S51, D2S1338, and FGA. The combined PD of the 15 loci was greater than 0.99999999999999999. This indicates that these 15 STRs in the Identifiler multiplex are highly discriminating, and as such would...
Table 3 Observed forensic efficiency parameters for 15-STR loci in Malay population

|        | D8S1179 | D21S11  | D7S820  | CSF1PO  | D3S1358 | TH01    | D13S317 | D16S539 | D2S1338 | D19S433 | vWA     | TPOX    | D18S51  | D5S818  | FGA     |
|--------|---------|---------|---------|---------|----------|---------|----------|----------|----------|----------|---------|---------|----------|---------|---------|
| p-value| 0.3517  | 0.6331  | 0.6987  | 0.9235  | 0.7272   | 0.0668  | 0.8563   | 0.1829   | 0.3124   | 0.7497   | 0.1002  | 0.4541  | 0.2508  | 0.9045  | 0.871   |
| PD     | 0.957   | 0.957   | 0.918   | 0.869   | 0.889    | 0.914   | 0.923    | 0.920    | 0.970    | 0.940    | 0.930   | 0.793   | 0.962   | 0.915   | 0.971   |
| PE     | 0.639   | 0.690   | 0.531   | 0.434   | 0.501    | 0.511   | 0.606    | 0.579    | 0.667    | 0.600    | 0.611   | 0.337   | 0.690   | 0.573   | 0.754   |
| TPI    | 2.80    | 3.28    | 2.10    | 1.69    | 1.96     | 2.01    | 2.54     | 2.37     | 3.04     | 2.51     | 2.58    | 1.38    | 3.28    | 2.34    | 4.16    |
| PIC    | 0.83    | 0.83    | 0.75    | 0.66    | 0.70     | 0.74    | 0.76     | 0.76     | 0.86     | 0.79     | 0.77    | 0.57    | 0.84    | 0.74    | 0.87    |
| HT     | 82.1%   | 84.8%   | 76.2%   | 70.4%   | 74.5%    | 75.1%   | 80.4%    | 78.9%    | 83.6%    | 80.1%    | 80.6%   | 63.6%   | 84.8%   | 78.6%   | 88.0%   |

Total Alleles 682 682 682 682 682 682 682 682 682 682 682 682 682 682 682

p-value: Possible deviations from H-W equilibrium. PD: Power of Discrimination. PE: Power of Exclusion. TPI: Typical Paternity Index. PIC: polymorphic information contents. HT: Heterozygotes.
be applicable to forensic casework in the Malaysian population.

The allele frequencies in the Malay population were compared with those in Chinese, Iranian, Belgian, and African populations using Fisher’s exact test. The probability values obtained are shown in Table 4.

The average coefficient of correlation was strongest in the order of Africa (0.092522), Belgium (0.264822), Iran (0.404363), and China (0.706661). The following 10 STRs showed the strongest correlation with Chinese ethnicity: D8S1179, D21S11, D7S820, TH01, D13S317, D2S1338, D19S433, vWA, D18S51, and D5S818. Meanwhile, CSF1PO, D3S1358, and FGA showed a close relationship with Iran, TPOX with Belgium, and D16S539 with Africa. These results indicated that Chinese ethnicity was the most closely related to Malay, while African was the least. This is consistent with what is known about the geographical conditions and history of prehistoric human migration in this region. Some of the STRs investigated (TH01, D5S818, D13S317, and D21S11) showed only a weak correlation with other ethnicities, suggesting that they would be useful for purposes of discrimination, particularly TH01. On the other hand, FGA and CSF1PO were unusual in that they showed almost no ethnic difference.

Such data on allele frequency are applicable to the statistical evaluation of DNA evidence in human identification, as well as to the determination of ethnicity in healthy populations. Further study is needed, however, on STR data from the Y and X chromosomes of these Malaysian samples. This group will also investigate STRs in Japanese samples and compare them with those from other populations, including those of the American continent.

Table 4  Probability values obtained from differentiation tests comparing Malaysian, with Chinese\(^2,5\), Iranian\(^7\), Belgian\(^3\), and African\(^1\) populations

| References | China     | Iran      | Belgium | Africa  |
|------------|-----------|-----------|---------|---------|
| D8S1179    | 0.925518  | 0.374239  | 0.72343 | 0.047934 |
| D21S11     | 0.61059   | 0.16384   | 0.27289 | 4.79E-05 |
| D7S820     | 0.80032   | 0.72918   | 0.29042 | 0.0063676|
| CSF1PO     | 0.65948   | 0.95779   | 0.69937 | 0.0458882|
| D3S1358    | 0.79138   | 0.84468   | 0.0746  | 0.229334 |
| TH01       | 0.03437   | 0.03208   | 0.00012 | 0.0007841|
| D13S317    | 0.86858   | 0.05724   | 0.00614 | 7.78E-12 |
| D16S539    | 0.59028   | 0.16034   | 0.0081  | 0.623683 |
| D2S1338    | 0.99131   | 0.33926   | 0.05866 | 0.0063498|
| D19S433    | 0.87137   | 0.31704   | 0.00071 | 0.0924051|
| vWA        | 0.82365   | 0.26104   | 0.4362  | 2.69E-05 |
| TPOX       | 0.41493   | 0.29678   | 0.64312 | 6.87E-05 |
| D18S51     | 0.67504   | 0.51755   | 0.48713 | 0.023826 |
| D5S818     | 0.85403   | 0.01664   | 0.00017 | 8.12E-05 |
| FGA        | 0.90793   | 0.96762   | 0.36127 | 0.422423 |
| **Average**| **0.70666** | **0.40436** | **0.26482** | **0.0925224** |
References

1) Alves C, Gusmão L, Damasceno A, Soares B, Amorim A (2004) Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 System) and a report on genotypic variations. Forensic Sci Int 139: 201–205.

2) Chen L, Lu H, Qiu P, Yang X, Liu C (2015) Polymorphism analysis of 15 STR loci in a large sample of Guangdong (Southern China) Han population. Legal Medicine 17:489–492.

3) Decorte R, Engelen M, Larno L, Nelissen K, Gilissen A, Cassiman JJ (2004) Belgian population data for 15 STR loci (AmpFlSTR® SGM Plus and AmpFlSTR™ profiler PCR amplification kit). Forensic Sci Int 139:211–213.

4) Nakamura Y, Minaguchi K (2010) Sixteen X-chromosomal STRs in two octaplex PCRs in Japanese population and development of 15-locus multiplex PCR system. Int J Legal Med 124:405–414.

5) Nata M, Kimura T, Hashiyada M, He P, Yan W, Li X, Funayama M, Sagisaka K (1999) Allele frequencies of eight STRs in Japanese and Chinese. Int J Legal Med 112:396–399.

6) Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86: 248–249.

7) Shepard EM, Herrera RJ (2006) Iranian STR variation at the fringes of biogeographical demarcation. Forensic Sci Int 158:140–148.

8) Tereba A (1995) Tools for analysis of population statistics. Profiles DNA 2:14–16.

Correspondence:
Dr. Yasutaka Nakamura
Department of Forensic Odontology and Forensic Anthropology,
Tokyo Dental College,
2-9-18 Misaki-cho, Chiyoda-ku,
Tokyo 101-0061, Japan
E-mail: ynakamura@tdc.ac.jp