

STR polymorphisms in Philippine ethnolinguistic groups: evaluation of forensic utility

Jasmin Jiji MIRANDA¹, Saturnina C. HALOS²

¹National Institute of Molecular Biology and Biotechnology, University of the Philippines, Diliman,
Quezon City, Philippines

²DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines, Diliman,
Quezon City, Philippines

Abstract. Population data was collected for the STR loci F13A01, FES/FPS, HUMvWA, and HUMTH01, in three major Philippine ethnolinguistic groups and used to estimate statistical parameters for identity testing in forensic work on Filipinos. The Cebuano, Ilocano, and Pampango populations in the Philippines were studied because they are among the biggest linguistic groups in the country, thus their genotypic profiles should substantially represent those of many Filipinos. The number of alleles varied from 4 to 9 at all loci, falling within the range observed for other local and world populations. Pairwise comparisons of the allele frequency distributions showed no statistical differences among the populations. The test for linkage equilibrium showed no evidence of non-random association of alleles across the physically unlinked loci in any of the three populations. The four loci combined gave an exclusion power of ≥ 0.9995 and a power of paternity exclusion of 0.8859-0.9389.

Key words: STR polymorphisms, forensics, genetic database, Philippines.

Introduction

The Philippines today has approximately 110 languages in the lexicostatistical sense of mutually unintelligible codes (MCFARLAND 1994). With the exception of imported languages, e.g. Chinese and Spanish, all these are generally agreed to belong to the Western Malayo-Polynesian branch of the Austronesian family of languages. The Philippine languages are further classified as Northern, Meso or Southern Philippine languages on the basis of geographical and ethnolinguistic

Received: August 18, 2003. Accepted: October 12, 2003.

Corresponding: J.J. MIRANDA, National Institute of Molecular Biology and Biotechnology, College of Science, University of the Philippines, Diliman, Quezon City 1101 Philippines, e-mail: jmiranda@pacific.net.sg

variations. The diversity of extant languages, the archipelagic geography of the country, and the differences in demographic profiles, offer interesting opportunities to explore genetic relations between ethnolinguistic populations in the country. To date, however, interpopulational genetic studies on Filipinos have so far been limited mainly to classic protein-based studies (OMOTO 1984, HORAI et al. 1981) and mitochondrial DNA polymorphism analysis (HARIHARA et al. 1988) on tribal groups which have a limited impact on the genetic configuration of modern urban Philippine populations on the whole; and STR analysis of regional populations defined by geographic and socio-political criteria (HALOS et al. 1998, 1999, TABBADA et al. 2002) and ethnically heterogeneous groups with either limited sample size (PARRA et al. 1999) or based on migrant individuals with undefined sampling criteria (PU et al. 1999). While those studies have helped to describe patterns of genetic variation among Filipinos and Asia-Pacific neighbours, there is still a general lack of genetic information for most Philippine populations, especially those defined primarily by language and ethnicity. In this study we constructed genetic databases of the short tandem repeat (STR) loci F13A01, FES/FPS, HUMvWA, and HUMTHO1, for three of the biggest linguistic groups in the Philippines: Cebuano, Ilocano and Pampango. The Cebuano population, located in the Visayas group of islands of central Philippines, is classified as a Meso-Philippine language subgroup and is geographically separated from the Northern Philippine subgroups Ilocano and Pampango of Luzon Island. The Cebuano, Ilocano and Pampango linguistic groups account for about 25%, 11% and 4%, respectively, of the Philippine population (GRIMES 1996). For forensic purposes, determining genetic variation among these populations would be helpful in assessing whether population databases should be ethnic-specific or aggregated, in view of reported differences in the probabilities of occurrence of a DNA profile across populations (GALLO et al. 1997) and nonhomogeneity among subpopulation regional databases (TABBADA et al. 2002). With this rationale, in this study we analysed the four STR loci in the three Philippine populations to assess the nature and degree of variation in these loci and to use the population data to estimate statistical parameters for identity testing in forensic work on Filipinos.

STR loci are now the preferred markers for human identification in forensics due to their high allelic diversity, abundance in the human genome, and amenability to PCR-based methods of analysis and automated detection technology. STR population genetic databases have been constructed worldwide (BUDOWLE 2001a, SUN et al. 2003) as a basis for calculating probabilities of a DNA match in the relevant population. The STR loci F13A01, FES/FPS, HUMvWA, and HUMTHO1, were chosen on the basis of their heterozygosity and validated use as forensic markers in other laboratories (HAMMOND et al. 1994, PESTONI et al. 1995), and availability of population data for comparison. In this study, the estimation of forensic efficiency values is based primarily on the National Research Council (NRC) Report (1996) entitled "The Evaluation

of Forensic DNA Evidence”, whose recommendations for such computations are becoming standard (SUN et al. 2003), and on leading references in the field (BRENNER, MORRIS 1990, HAMMOND et al. 1994).

Material and methods

DNA was isolated by using the standard proteinase K method from buccal/blood samples collected from unrelated donors belonging to three of the biggest Philippine ethnolinguistic groups: Cebuanos (in Cebu Island in the Visayas, Central Philippines), Ilocanos of Ilocos, and Pampangos of Tarlac (both in Luzon Island, Northern Philippines). Each population was represented by about 50 individuals. Only native speakers whose parents also belonged to the same language group and who originated from the area were included. PCR amplification at the F13AO1, FESFPS, HUMvWA, and HUMTHO1 loci, was done according to HALOS et al. (1999). Characteristics of the STR loci are shown in Table 1. The PCR products were electrophoresed alongside allelic ladders and fluorescently detected with the ALFExpress DNA Sequencer (Pharmacia Biotech). Post-run analysis and gene scoring were done by Allele Links ver. 1 software (Pharmacia Biotech).

Table 1. Characteristics of the STR loci F13AO1, FES/FPS, HUMvWA, and HUMTHO1

STR	Locus definition	Location	Repeat
F13AO1	human coagulation factor XIIIa subunit gene	6p24–25	AAAG
FES/FPS	human c-fes/fps proto-oncogene	15q25–qter	AAAT
HUMvWA	human von Willebrand factor gene	12p12–pter	AGAT
HUMTHO1	human tyrosine hydroxylase gene	11p15.5	AATG

The exact/probability test for the Hardy-Weinberg Equilibrium (HWE), estimation of expected heterozygosity with the use of Levene’s correction, and Fisher’s exact test for linkage equilibrium, were performed by Genepop 1.2 software (RAYMOND, ROUSSET 1995). The exclusion power of a locus, which represents the probability that two persons do not have the same genotype, was calculated as $1 - P_i$, and the combined exclusion power of the loci as $1 - (\prod P_i)$, where P_i is the sum of squares of the frequencies of all genotypes at a locus (NRC 1996). The power of paternity exclusion, PE , and PE_{typical} of n loci, defined as the fraction of individuals who would not have the same DNA profile presented in a paternity case, were calculated from the following formulas: $PE = h^2(1-2hH^2)$, where h is heterozygote frequency and H is homozygote frequency; and $PE_{\text{typical}} = 1 - \prod(1 - PE_i)$ (BRENNER, MORRIS 1990).

Results and discussion

Allele frequencies

Our study revealed genetic variation at four STR loci in three well-defined ethnolinguistic groups in the Philippines. Table 2 shows the summary statistics of the F13AO1, FES/FPS, HUMvWA, and HUMTHO1 STR loci, in the Cebuano, Ilocano, and Pampango populations. With respect to the distribution of alleles, each STR locus was found to be polymorphic in all populations. The number of alleles varied from 4 to 9 at all loci, falling within the range observed for other local and world populations (HALOS et al. 1999, PU et al. 1999, TABBADA et al. 2002, Genetic Identity Population Data, www.promega.com). As a rule of thumb, the higher the number of alleles at a locus, the higher the potential number of heterozygotes, and thus the more powerful this system can be for resolving small differences in the distribution of allele frequencies in the populations (WEIR 1996). In our study, up to several alleles were observed for each locus, especially at F13AO1 and HUMvWA, but only a few were relatively common. The other alleles were rare, e.g. alleles 14 and 16 at F13AO1 and alleles 12 and 13 at HUMvWA, thus contributing little to heterozygosity. Pairwise comparisons of the allele frequency distributions showed no statistical differences among the populations after Bonferroni correction. The overall similarities in allele frequencies suggest that the STR polymorphisms predate differentiation events, e.g. dispersals or mutation, in the three Philippine populations. Moreover, they support the lexico-statistical classification of these populations under the same Western Malayo-Polynesian ethnolinguistic branch (MCFARLAND 1994), reflecting shared ancestry. The similarity in heterozygosity levels (Table 2) among the geographically isolated Cebuano population and the geographical neighbours, Ilocano and Pampango populations, suggest that these modern urban populations have approached the same degree of genetic variability and/or drift-mutation and migration balance. However, comparisons with other Philippine populations (HALOS et al. 1999, PU et al. 1999, TABBADA et al. 2002) showed that the Pampango population is significantly different ($p = 0.002$) from the Visayas (Region VIII) (TABBADA et al. 2002). As these populations are geographically separated, i.e. the Pampango in northern Philippines and the Visayas in central Philippines, the distance, and linguistic and demographical differences between these populations may have helped to set genetic boundaries. It should be noted that sampling criteria in those Philippine studies differed as well, as TABBADA et al. (2002) used samples from blood repositories in the region of interest, while we used language and residence as criteria. Expectedly, the three populations in this study are distinguishable from a non-Asian (Caucasian) population (Genetic Identity Population Data, www.promega.com) at all loci, except at HUMvWA (data not shown). The occurrence of genetic variation among the populations emphasizes the need to construct and use appropriate databases for forensic applications, as already recommended previously (GALLO et al. 1997, TABBADA et al. 2002).

Table 2. Allele frequencies and forensic statistics of four STR loci in three Philippine ethnolinguistic populations

Allele	F13AO1			Allele	FES/FPS			Allele	HUMvWA			Allele	HUMTHO1		
	Cebuano (n=50)	Ilocano (n=50)	Pampango (n=48)		Cebuano (n=49)	Ilocano (n=50)	Pampango (n=50)		Cebuano (n=48)	Ilocano (n=50)	Pampango (n=50)		Cebuano (n=51)	Ilocano (n=50)	Pampango (n=48)
3.2	0.180	0.260	0.281	8			0.010	12	0.010			6	0.137	0.080	0.156
4	0.060	0.050	0.063	10	0.031	0.060	0.020	13	0.010			7	0.343	0.280	0.250
5	0.130	0.110	0.094	11	0.429	0.320	0.380	14	0.094	0.130	0.220	8	0.127	0.070	0.063
6	0.600	0.580	0.500	12	0.296	0.420	0.400	15	0.146	0.130	0.040	9	0.353	0.430	0.396
7	0.010		0.021	13	0.214	0.200	0.190	16	0.125	0.200	0.160	9.3		0.090	0.083
8			0.021	14	0.031			17	0.281	0.210	0.140	10	0.039	0.050	0.052
10	0.010							18	0.198	0.220	0.290				
11	0.010							19	0.115	0.110	0.140				
14			0.010					20	0.021		0.010				
16			0.010												
h^a	0.620	0.580	0.740		0.816	0.720	0.653		0.771	0.820	0.720		0.726	0.620	0.646
H^b	0.380	0.420	0.260		0.184	0.280	0.347		0.229	0.180	0.280		0.274	0.380	0.354
HWE^c	0.9147	0.9230	0.3288		0.0637	0.8758	0.7149		0.3351	0.9903	0.3599		0.7035	0.1266	0.1733
EP^d	0.7897	0.7670	0.8333		0.8390	0.8343	0.8139		0.9447	0.9424	0.9311		0.8736	0.8773	0.7670
Power of PE^e	0.3156	0.7324	0.5072		0.6296	0.4599	0.3595		0.5460	0.6367	0.4599		0.4688	0.3156	0.3495

h^a heterozygote frequency, H^b homozygote frequency, c exact test p value (Genepop 1.2, RAYMOND M., ROUSSET F. 1995); d exclusion power of the locus (NRC, 1996); e power of paternity exclusion (BRENNER C., MORRIS J. 1990)

Some DNA samples failed to amplify at some loci, most probably due to mutations at the primer-binding site. Notwithstanding, the overall validity of the data for forensic applications is not affected by such a low incidence of null alleles (BUDOWLE et al. 2001b, SUN et al. 2003). Inasmuch as the present study was based on gel migration differences of the alleles, the contribution of single nucleotide polymorphisms to allelic diversity could not be ascertained but previous studies have already demonstrated the fidelity of most STRs, with known variants well documented (URQUHART et al. 1994, PESTONI et al. 1995). Further, the sensitivity, high level of precision, and accuracy of the automated detection system used in this study ensure the overall reliability of results.

For forensic purposes, HWE conformity is desired in the estimation of statistical parameters, as the calculations rely heavily on the condition of random genetic assortment. The allele frequencies at all four loci in all the populations were found to be in the HWE (Table 2). The allele frequencies in the HWE can thus be used to determine the genotype frequencies in the true population and this comes in handy when the genotype of a DNA sample has not been observed in the relevant population. The test for linkage equilibrium showed no evidence of non-random association of alleles across the physically unlinked loci in any of the three populations. Therefore, we could multiply genotypic frequencies across the loci F13AO1, FES/FPS, HUMvWA, and HUMTHO1, to obtain the DNA profile frequency (product rule) in each population. This operation is helpful in calculating the random-match probability of a locus, in cases when there is a need to assess the chances that someone in the population, other than the suspect, could have the same DNA profile (NRC, 1996).

Exclusion power of a locus

Since the frequency of a DNA profile is already a composite of individual genotype frequencies at different loci, computations must start at the single locus level. The probability that two randomly chosen persons have the same unspecified genotype at a locus is the sum of squares of the frequencies of all genotypes at that locus (NRC, 1996). This probability is designated as P_i , or the individualization potential of a locus (HAMMOND et al. 1994), and is used to calculate the exclusion power of a locus, $1 - P_i$. The higher the exclusion power of a locus, the more efficient it is in discriminating between members of the population. Of the four loci, HUMvWA consistently showed the highest exclusion power (> 0.93) in the three population samples, while F13AO1 showed the lowest in Ilocano and Cebuano, and HUMTHO1, in Pampango (Table 2).

The four loci combined gave an exclusion power of 0.9998, 0.9997, and 0.9995, for the Cebuano, Ilocano, and Pampango population samples, respectively. This means that when used together, these loci can distinguish samples from different individuals with a probability of 99.98%, 99.97%, and 99.95%, respectively, thus clearing many innocent suspects of charges.

Power of paternity exclusion

Another parameter used to evaluate the strength of a locus to exclude falsely accused individuals is the power of paternity exclusion, or *PE*. It represents the percentage of individuals in the relevant population who would not share the same DNA profile presented in a paternity case (BRENNER, MORRIS 1990). The higher the *PE* value, the more non-fathers are excluded.

Single-locus *PE* values ranged from 0.3156 to 0.7324 (Table 2), which indicated low degrees of exclusionary power for the loci when used individually. However, combined values for the four loci demonstrated their forensic utility, especially for multilocus-based analysis with additional loci to increase the power of the system. The Cebuano sample gave the highest PE_{typical} of 0.9389, followed by Ilocano with 0.9016, and Pampango with 0.8859. The differences in the paternity statistics among the populations further emphasizes the importance of using appropriate databases to estimate statistical parameters for forensic applications.

In summary, we report here a 4-locus STR database in a set of ethnolinguistically defined Philippine populations. The similarities in the population data should be taken with caution in forensic applications as the frequency of the same DNA profile may differ across populations. The benefit of using the most relevant population data to estimate forensic parameters cannot be overstated in terms of lessening the chances of false convictions and mistaken paternity. Population genetic characteristics desired in forensic analysis, such as conformity to HWE and linkage equilibrium across loci, were satisfied by the four loci in the Philippine populations. The present database thus validates the use of the four STR loci for forensic applications. However, as expanded multilocus tests (e.g. 7- to 16-loci multiplexes) are becoming standard in forensic laboratories worldwide, supplementation of the present set with additional informative loci is highly recommended for identity testing in these Philippine populations. Although many Philippine populations and genetic loci remain to be studied, our results provide important information on genetic variation among Filipinos and contribute to understanding global genetic variation.

REFERENCES

- BRENNER C., MORRIS J. (1990). Paternity index calculations in single locus hypervariable DNA probes: Validation and other studies. Proc. Int. Symp. Human Identification. Promega Corporation: 21-53.
- BUDOWLE B., SHEA B., NIEZGODA S., CHAKRABORTY R. (2001). CODIS STR loci data from 41 sample populations. J. Forensic Sci. 46(3): 453-489.
- BUDOWLE B., MASIBAY A., ANDERSON S.J., BAMA C., BIEGA L., BRENNEKE S. et al. (2001). STR primer concordance study. Forensic Sci. Int. 124: 47-54.
- GALLO J.C., THOMAS E., NOVICK G.E., HERRERA R.J. (1997). Effects of subpopulation structure on probability calculations of DNA profiles from forensic PCR analysis. Genetics 101(1): 1-12.
- GENETIC IDENTITY POPULATION DATA. www.promega.com
- GRIMES B.F. (ed.) (1996). Ethnologue: Languages of the world, 13th edn. Summer Institute of Linguistics, Inc. Dallas, USA.

- HORAI S., OMOTO K., JUJI T., SONOZAKI H., MITSUI H., MISAWA S., SUMPAICO J., MERCADO A. (1981). The HLA antigens of 2 Negrito populations in the Philippines. *Tissue Antigens* 17(3): 343-348.
- HALOS S.C., FORTUNO III E.S., FERREON A.C.M., CHU J.Y., MIRANDA J., HARADA S., BENECKE M. (1998). Allele frequency distributions of the polymorphic STR loci HUMVWA, HUMFES, HUMF13A01 and the VNTR D1S80 in a Filipino population from Metro Manila. *Int. J. Legal. Med.* 111: 224-226.
- HALOS S.C., CHU J.Y., FERREON A.C.M., MAGNO M.M.F. (1999). Philippine population database at nine microsatellite loci for forensic and paternity applications. *Forensic Sci. Int.* 101: 27-32.
- HARIHARA S., SAITOU N., HIRAI M., GOJOBORI T., PARK K.S., MISAWA S., ELLEPODA S.B., ISHIDA T., OMOTO K. (1988). Mitochondrial DNA polymorphism among five Asian populations. *Am. J. Hum. Genet.* 43: 134-143.
- HAMMOND H., JIN L., ZHONG Y., CASKEY T., CHAKRABORTY R. (1994). Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* 55: 175-189.
- MCFARLAND C. (1994). Subgrouping and number of the Philippine languages or: How many Philippine languages are there? *Philipp. J. Linguistics* 25(1-2): 75-84.
- NATIONAL RESEARCH COUNCIL (1996). The evaluation of forensic DNA evidence. National Academic Press, Washington DC.
- OMOTO K. (1984). The Negritos: Genetic origins and microevolution. In: *Out of Asia: Peopling the Americas and the Pacific* (Kirk R., Szathmary E., eds.). Australian National University Printery: 123-132.
- PARRA E., SAHA N., SOEMANTRI A.G., MCGARVEY S.T., HUNDREISER J., SHRIVER M.D., DEKA R. (1999). Genetic variation at 9 autosomal microsatellite loci in Asian and Pacific populations. *Hum. Biol.* 71: 757-779.
- PESTONI C., LAREU M.V., RODRIGUEZ M.S., MUNOZ I., BARROS F., CARRACEDO A. (1995). The use of the STRs HUMTHO1, HUMvWA31/A, HUMF13A1, HUMFES/FPS, HUMLPL in forensic application: validation studies and population data for Galicia (NW Spain). *Int. J. Legal. Med.* 107: 283-290.
- PU C.E., HSIEH C.M., CHEN M.Y., WUF C., SUN C.F. (1999). Genetic variation at nine STR loci in populations from the Philippines and Thailand living in Taiwan. *Forensic Sci. Int.* 106: 1-6.
- RAYMOND M., ROUSSET F. (1995). Genepop (version 1.2): A population genetic software for exact tests and ecumenicism. *J. Hered.* 86: 248-249.
- SUN G., MCGARVEY S.T., BAYOUMI R., MULLIGAN C.J., BARRANTES R., RASKIN S., ZHONG Y., AKEY J., CHAKRABORTY R., DEKA R. (2003). Global genetic variation at nine short tandem repeat loci and implications on forensic genetics. *Eur. J. Hum. Genet.* 11: 39-49.
- TABBADA K.A., MAGNO M.M.F., DELFIN F.C., CALACAL G.C., TAN M., FERREON J.C., HALOS S.C., DE UNGRIA M.C.A. (2002). Allele frequencies of eight short tandem repeat loci in three Visaya regional populations in the Philippines. *J. Forensic Sci.* 47(6): 1397-1398.
- URQUHART A., KIMPTON C.P., DOWNES T.J., GILL P. (1994). Variation in short tandem repeat sequences: a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. Leg. Med.* 107: 13-20.
- WEIR B.S. (1996). *Genetic data analysis II*. Sinauer Associates Inc. Publishers, Sunderland, Massachusetts.