

Characterization of the Philippine Y STR Population Involving 23 Loci:
Population Information, Mutation Analysis and Forensic Assessment
(1st Draft)

A Thesis
Presented to
The Faculty of Biology Department
De La Salle University
Taft Avenue, Manila

In Partial Fulfillment
of the Requirements for the Degree
of Master Science in Biology

by
Sheila Marie T. Angustia

October 2013

APPROVAL SHEET

The thesis proposal entitled **Characterization of the Philippine Y STR Population Involving 23 Loci: Population Information, Mutation Analysis and Forensic Assessment**" in partial fulfillment of the requirement for the Degree of Master of Science in Biology has been examined and is recommended for acceptance and approval.

Dr. Ma. Carmen Ablan-Lagman

Date

(Adviser)

The thesis proposal is accepted in partial fulfillment of the requirement for the Degree of Master of Science in Biology.

Chair

Date

ABSTRACT

Y-chromosomal STR markers have shown to be advantageous in analyzing male DNA in mixtures containing high female DNA fraction and as supplementary markers for paternity testing. To be useful in the Philippine population markers must first be evaluated for its variability and mutability. In this study, twenty-three Y-chromosomal STR markers were used in characterizing the Philippine male population. Haplotype and allele frequencies were determined in 629 unrelated Filipino males while mutation data were gathered from the analysis of 133 confirmed father and son transmissions. A haplotype diversity of 0.99992 with a standard error of 0.00230 was obtained. This defines the presence of 612 haplotypes from the sample of 629 individuals. Out of the 612 haplotypes, 596 haplotypes were found to be unique, 15 haplotypes were found to be shared by two individuals and 1 haplotype was shared by three individuals. The discrimination capacity was calculated to be 97.30%. Among the 3,059 meiotic transfers covered, 11 mutations were recorded with an average overall mutation rate of 3.59×10^{-3} in the population. Five mutations resulted in a gain of repeat and 6 resulted in a loss of repeat suggest the absence of neither expansion nor contraction bias. DYS635 locus with three mutational events has the highest mutational rate (2.26×10^{-2} (95% CI (4.7-64.5) $\times 10^{-3}$) across all loci. A haplotype with a duplicated allele in both father and son were observed at the DYS576 locus while an individual's haplotype was found to contain four alleles at DYS385 A/B. In addition a variant of DYS456 was observed to fall outside the marker range and a microvariant was observed at DYS576. Furthermore, information obtained from this research verified the usefulness and informative power of having additional markers for routine casework while concordance studies shows that both multiplex systems can be used to successfully type profiles and that the results of the Y Filer kit can be compared directly with the results of PPY23.

CHAPTER I

INTRODUCTION

Forensic DNA analysis is one of the most demanding investigative tools being used due to its social relevance. Advancements in this field were the outcome of the continuous development of functional methods in molecular biology. While autosomal microsatellite analysis serves as the most popular technique, other molecular techniques such as Y chromosomal microsatellite (Y STR), mitochondrial DNA (mtDNA) and single nucleotide polymorphism (SNPs) analyses are being explored. In line with these developments, forensic laboratories are being encouraged to use new techniques that would aid them in their quest for truth and justice.

Currently, DNA Analysis Branch of the PNP Crime Laboratory uses twenty (20) autosomal STR markers (D3S1358, D1S1656, D6S1043, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, vWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433 and FGA) for its routine forensic casework. However, due to variable nature of forensic evidence as well as the limitation set by autosomal STR technique in sexual assault cases involving mixed DNA profiles as well as deficient kinship investigation, the laboratory opted the inclusion of other forensic DNA method particularly Y chromosomal microsatellite (Y STRs) analysis in routine casework typing.

Goes et al (2005) mentioned that Y STR haplotype frequency data gathering in different population is foremost when evaluating genetic profile matches while mutation rates are fundamental in kinship examinations. Several studies concerning Y STR was done in the Philippines. Initially, a Y STR reference population database involving the European Minimal Haplotype (DYS19, DYS385 A/B, DYS 390, DYS 391, DYS392 and DYS 393) with a

population size of 106 from the National Capital region was reported by Miranda et al (2001). The database was then expanded by Tan et al (2003) by adding 105 additional profiles thereby increasing the population size to 211 using the same core loci set. Studies relating to the usefulness of Y STR in detecting male profiles to children who are victims of sexual abuse have emphasized the forensic relevance of Y STR (Maiquilla et al, 2011 and Delfin et al, 2005). Mutation studies have not been reported.

As forensic practitioners, a method that is effective, robust and reliable is preferred considering the voluminous cases being handled. Multiplex systems are being favoured due to ease of process and speed it offers.

Objectives of the Study

The goal of this research was to type 23 Y STR loci using the two latest commercially available multiplex systems (Y Filer by Applied Biosystems and Powerplex Y23 by Promega) in order to establish a more comprehensive reference population database for the Philippine population.

Specifically, this study aims to evaluate biostatistical population parameters (allele frequency, individual locus genetic diversity, haplotype diversity and discrimination capacity) from Philippine population database and compare the data from the previously reported NCR population database (Tan et al, 2002) to find out the effects of having an increased population size as well as sampling area. In addition, this study also intends to assess the forensic relevance of extending the number of loci as well as to evaluate the usefulness of six additional markers included in the newly released 23 Y STR system set.

Due to the crucial impact of mutation on kinship analysis, father and son transmissions were also examined to determine the overall mutation rate in the Philippine population as well as to identify the locus that exhibits the highest mutation rate. Further, this study aims to report the existence of non uniform alleles which includes duplications, deleted and null alleles. Lastly, two widely used commercial Y STR multiplex systems were evaluated for concordance.

Scope and Limitations

The current study will be limited to the analysis of 23 Y- chromosomal STR markers (DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456 and Y-GATA-H4) in the Philippine population. DNA profiles of 629 unrelated Filipinos was used for population studies while 133 father and son transmissions was in mutational analysis.DNA typing process is limited to use of two latest commercially available multiplex systems (Y Filer by Applied Biosystems and Powerplex Y23 by Promega) following the protocols established in the Philippine National Police Crime Laboratory. Confirmation of mutation thru sequencing will be limited to DYS635, DYS456 and DYS458 only.

CHAPTER II

REVIEW OF RELATED LITERATURE

Y chromosome is one of the smaller human chromosome with an estimated average size of 60 million base pairs (Mb). The heterochromatic region in the Y chromosome is located in the distal long arm (Yq) and constitutes more than half of the chromosome in some normal males but is virtually undetectable in others (Gusmao *et. al* 1999). Short tandem repeat polymorphisms on the non recombining part of the Y chromosome are inherited in the form of compound Y haplotypes and can only evolve by mutation (Brinkmann, 1999).

Y-chromosomal polymorphic markers have been extensively investigated in forensic medicine and have been recognized as a powerful tool for analyzing mixed forensic stains and for paternity testing (Rodig *et. al* 2007, Goes *et. al* , 2005). These markers are inherited along a male lineage, and are characterized by a moderate number of polymorphic loci.

In forensic genetics, Y chromosomal STR markers are typically used for offender identification, patrilineage tracking and victim recognition during mass disasters. Typically, sexual assault evidence such as vaginal swabs and underwear stains contain DNA admixtures containing low fraction of male DNA versus female DNA. On these samples, male DNA is separated from the female DNA thru differential extraction before analyzing using autosomal STR. Yet, occasionally, high concentrations of female DNA still exist even after the differential process. At these instances, amplification of the female DNA fraction is being favored than the

male fraction which makes the analysis of the male DNA to be difficult when analyzed with autosomal STR markers. Conversely, the male specificity of Y chromosomal STR markers renders its significance as supplementary markers relevant to processing of sexual assault evidence containing admixtures (Wurmb-Schwark *et.al* 2003 and Betz et al., 2001). Because Y STR markers target only the Y region of the male DNA, it will not amplify the female DNA and masking is prevented. Therefore, the profile generated will only come from the male assailant and analysis will be simpler. A study done by Johns et al (2006) showed that in cases of sexual assaults with a male perpetrator and a female victim, Y-STRs improve the chances of detecting low levels of the perpetrators DNA in a high background of the female victim's DNA.

Y chromosome STR has been shown to be a powerful tool in processing cases involving azoospermic semen (Dekairelle and Hoste 2000). Studies have demonstrated that Y STR analysis was successful in elucidating male profile (48%) despite a negative result for autosomal analysis depending on the amount of DNA available and the quality of DNA extracted (Dekairelle and Hoste 2000).

Application of Y chromosomal STRs to forensic casework becomes more informative than autosomal STR typing (plus amelogenin) alone especially if male suspect is unknown or when mixed stains from more than one male have to be investigated (Wurmb-Schwark *et.al* 2003). In forensic cases involving multiple male aggressors, autosomal STR analysis often provide inconclusive results while Y STR markers can be used to estimate the number of donors based on the presence of additional alleles present in the Y chromosome profile.

In addition, owing to its paternal inheritance, all paternally related males will have the same Y STR profile provided the absence of mutation. This characteristic can appropriately be used in deficiency paternity analysis. For example, if the DNA profile of the direct ascendant

(parent) or the direct descendants (child) of an unidentified person is not around, the DNA profile of a paternally related male like his brother can be used to identify him. Lastly, in cases of mass disasters, victims who are paternally related can easily be identified when processed using the Y STR (Goes *et. al*, 2005, Corach *et. al*, 2001).

A set of Y chromosome short tandem repeat (Y-STR) loci have been selected for use in human identification applications. Y-STR loci recommended by the SWGDAM (Scientific Working Group on DNA Analysis Methods) Y chromosome subcommittee in 2003 include the following: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, and the multi-copy locus DYS385 a/b (Butler, 2006). At present, several commercial kits are available which allows the ease and speed of typing (Thompson, 2013).

Determination of Y STR haplotype frequencies in different populations is highly significant for correct interpretation of genetic profile matches in paternity and forensic casework (Goes *et. al*, 2005, Goedbloed *et. al*, 2009). Occurrence of population substructure renders the need for microgeographic sampling when calculating for match as well as paternity probabilities. Thus, Y STR markers included in forensic casework has to be evaluated for its variability and usefulness in the population for which they will be used by obtaining haplotype diversities and understanding population structure.

In situations where direct comparison between evidence and a suspect is being made, mutation rates are not important. However, spontaneous mutation in the germline between putative father at any locus in kinship analysis may be misinterpreted due to differences in father and son pairs (Lee *et. al*, 2009). Thus, reliable mutability knowledge is needed to carefully choose loci to minimize error rate in kinship analysis and sample identification while obtaining maximum discriminatory power (Burgarella *et. al*, 2011).

Many regions of Y-chromosome are duplicated or even triplicated in some individuals thus complicating the potential mixture interpretation where multiple peaks can be interpreted as mixed profiles (Vieira *et. al.*, 2008). Thus, in addition to haplotype frequencies estimates of mutation rates are essential in making decisions concerning paternity testing and forensic casework. The more accurate method of estimating Y STR mutation rates is the direct observation of transmission between father and son as long as large numbers of genetic transfers (meioses) are covered by testing a large number of father and son pairs (Ballantyne *et. al.*, 2010). The proportion of allele mismatches in father-son transmission is currently the most widely used approach to obtain mutation rates for YSTRs. Owing to the low values of mutation rates, large number of father and son pairs must be genotyped to obtain accurate estimates (Burgarella *et al.*, 2011).

CHAPTER III

MATERIALS AND METHOD

DNA Samples

Two hundred thirty human male DNA samples (97 fathers and 133 sons) from previous paternity cases and 532 male reference standards from previous forensic cross matching examinations submitted to the Philippine National Police Crime Laboratory – DNA Analysis Branch (PNPCL-DAB) were collated for DNA profiling using Y STR markers. Human sample collection followed the guidelines outlined by the PNP Crime Laboratory Standard Operational Procedures.

DNA Extraction

Genomic DNA samples were extracted using Chelex, Purelink and Promega DNA IQ™ System following manufacturers' protocol and quantified using Thermo Scientific NanoDrop spectrophotometer.

PCR Amplification

DNA samples were using amplified using *PowerPlex® Y23* System PCR Amplification and AmpFlSTR® *Yfiler®* PCR Amplification kit with some modifications in the recommended manufacturer's protocol. Applied Biosystems Veriti™ thermal cycler were used in all amplification reactions.

PowerPlex® Y23 System PCR Amplification

DNA samples were amplified using *PowerPlex® Y23* System to a total volume of 6.25uL amplification reaction under the following cycling conditions: initial denaturation of 96°C for 2

min, 30 cycles of 94°C for 10s, 61°C for 1 min and 72°C for 30s. This will be followed by 20min hold at 60°C and final soak at 4°C for 5 mins.

AmpFISTR® Yfiler® PCR Amplification

DNA samples were amplified using AmpFISTR® Yfiler® kit to a total volume of 6.25uL amplification reaction under the following cycling conditions: initial denaturation of 95°C for 11 min; 30 cycles of 94°C for 1 min, 61°C for 1 min and 72°C for 30s; 80min hold at 60°C and final soak at 4°C for 5 mins. Each amplification reaction contains 2.3 μL AmpFISTR® PCR Reaction Mix, 1.25 μL AmpFISTR® Yfiler® Primer Set, 0.2 μL AmpliTaq Gold® DNA Polymerase, 1.25 μL molecular grade H₂O DNA and 1.0 μL template DNA.

DNA typing

PowerPlex® Y23 System Typing Procedure

Separation of PCR fragments were attained using the Applied Biosystems 3500 Genetic Analyzer with a 36 cm capillary array and POP-4® Polymer (Life Technologies™) set on the following injection parameters: 1.2kv, 24s injection time. Each analysis reaction was prepared by adding 10 μL *Hi-Di™ Formamide* (Life Technologies™), 0.8 μL size standard and 1 μL PCR product. Samples were denatured at 95°C for 3 minutes and snap cooled in an ice bath prior to injection. Data were gathered using *3500 Data Collection Software v1.0*. Allele sizing and calling were carried out on the *GeneMapper® ID-X v1.2* (Life Technologies™) using the PowerPlex® Y23 bins and text files v1.0. The allelic ladder mix provided on the kit together with the CC5 Internal Lane Standard 500 Y23 (ILS) was utilized during analysis. 2.3.2

AmpFISTR® Yfiler® Typing Procedure

Separation of PCR fragments were attained using the Applied Biosystems 3500 Genetic Analyzer with a 36 cm capillary array and POP-4® Polymer (Life Technologies™) under the following injection parameters: 1.2kv, 15s injection time. Each analysis reaction was prepared by adding 8.7 μ L *Hi-Di™ Formamide* (Life Technologies™), 0.3 μ L GeneScan™ 600 LIZ® Size Standard and 1 μ L *PCR product*. The *AmpFISTR® Yfiler®* allelic ladder mix provided on the kit was utilized during analysis. Samples were denatured at 95°C for 5 minutes and snap cooled in an ice bath for another 5 minutes prior to injection. Data were gathered using 3500 Data Collection Software v1.0. Allele sizing and calling were carried out on the *GeneMapper® ID-X* v1.2 (Life Technologies™) using the *AmpFISTR® Yfiler®* bins and text files v2.

Quality Control

The laboratory undergoes internal lane standard testing.

Population Studies

Six hundred twenty nine Filipino male reference samples (97 DNA samples from the fathers obtained from paternity cases and 532 male reference standards) taken from previous routine forensic cases of Philippine National Police Crime Laboratory DNA Analysis Branch were processed using two different commercial kits. Samples originated from fourteen different regions of the country. The number of samples obtained per region is presented in table 1. No connections or relationship between the subjects were known.

Table 1. Population Samples Obtained Per Region

REGION	NUMBER OF SAMPLES	REGION	NUMBER OF SAMPLES
NCR	266	Region 8	10
Region 1	50	Region 9	6
Region 2	12	Region 10	9
Region 3	46	Region 11	7
Region 4a	72	Region 12	5
Region 4b	12	Region 13	14
Region 5	45	CAR	38
Region 6	7	ARMM	3
Region 7	7	Total = 629	

Analysis of Data

Analysis of Data for Population Studies

Allele frequencies were obtained by gene counting while haplotype frequencies were determined using Arlequin Software V. 3.5.1.2. Gene and Haplotype diversity is computed according to Nei (1987) where x_i is the frequency of each haplotype (allele) in the sample and N is the sample size.

$$H = \frac{N}{N-1} \left(1 - \sum_i x_i^2\right)$$

Discrimination capacity is calculated by dividing the number of observed haplotypes by the number of sampled individuals.

Cumulative power of discrimination was computed by the following formula:

$$CPD = 1 - \{(1-P_1) \times (1-P_2) \times \dots \times (1-P_n)\}$$

Where P_1, P_2, \dots, P_n equals the gene diversity of specific loci.

Population Substructure was determined thru AMOVA using Arlequin Software V. 3.5.1.2.

Annotated recorded

Mutation Analysis

A total of 133 father to son transmission was screened for mutation. Two hundred thirty human male DNA samples (97 fathers and 133 sons) from previous paternity cases were included on this study.

CHAPTER V

RESULTS AND DISCUSSION

Population Homogeneity

In order to establish the Philippine population database, samples from different provinces of the Philippines were assembled and initially categorized into three regional groups: Luzon (N=556), Visayas (N=25) and Mindanao (N=48). The use of regional Y-STR haplotype database as recommended by the International Society of Forensic Genetics (ISFG) requires verification of the absence of population substructure before pooling populations. Whilst, previous study of the Philippine population using the autosomal STR reveals that population differentiation exist between regional samples (Maiquilla et al, 2011), population differentiation using the Y STR markers was assessed.

Regional sample databases were compared using AMOVA test and revealed that most molecular variation was due to variation within population (98.94%) rather than among them. In addition, AMOVA results (Figure 1) indicated there is no significant difference between Luzon, Visayas and Mindanao populations ($P=0.0671 > 0.05$). Furthermore, fixation index (RST like = 0.016 < 0.05) suggests that there is no population substructure and therefore samples can be pooled creating the Philippine population database.

Figure 1. Estimation of Population Differentiation Luzon, Visayas and Mindanao Populations based on Analysis of Molecular Variance (AMOVA)

Distance method: Sum of squared size difference (GST)

AMOVA design and results :

Weir, B.S. and Cockerham, C.C. 1984.
Excoffier, L., Smouse, P., and Quattro, J. 1992.
Weir, B. S., 1996.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	2	184.967	0.47632 Va	1.59
Within populations	1255	36805.787	29.32732 Vb	98.41
Total	1257	36990.754	29.80264	
Fixation Index FST :		0.01595		

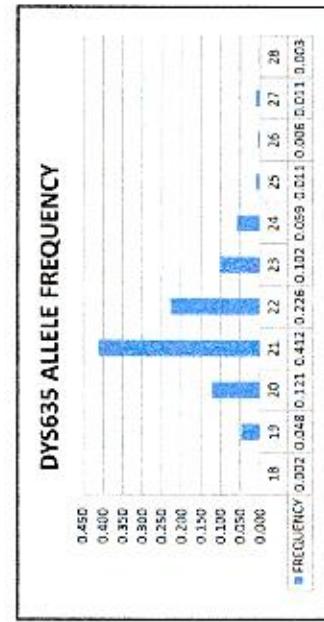
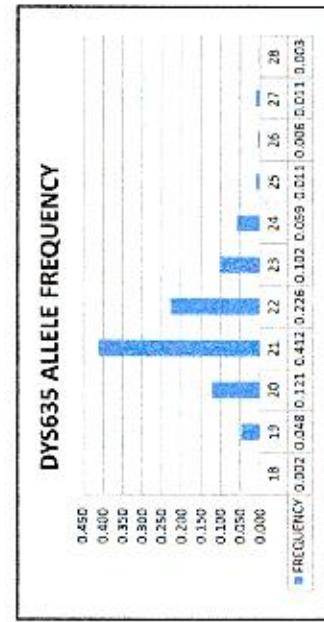
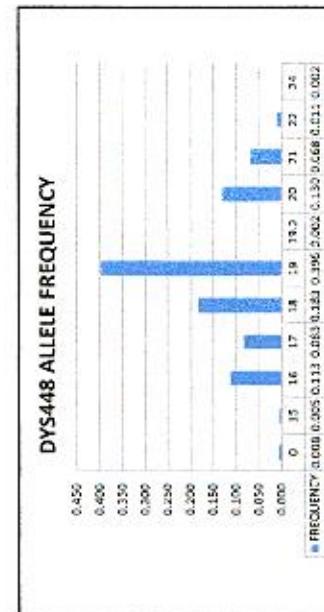
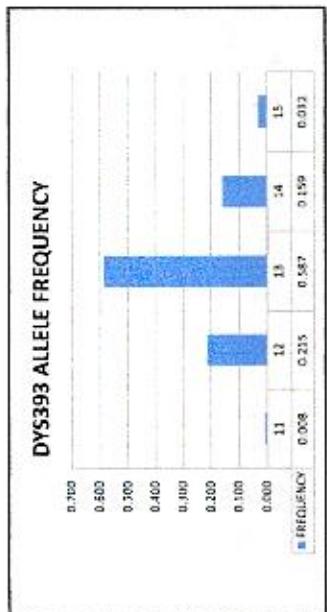
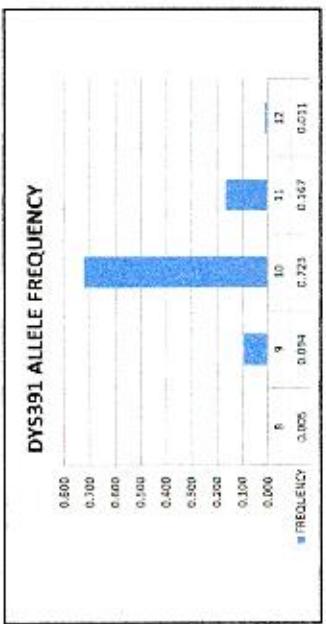
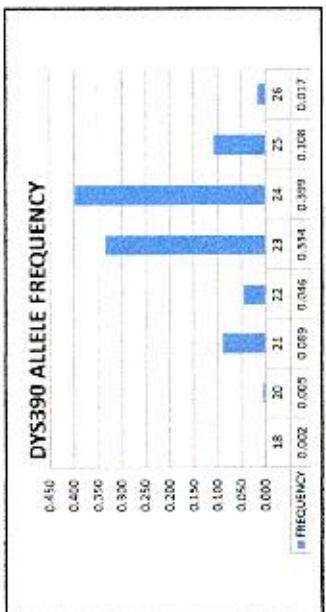
Significance tests (10100 permutations)

Va and FST : P(rand. value > obs. value) = 0.06713
P(rand. value = obs. value) = 0.00000
P-value = 0.06713±0.00221

Single Locus Analysis

Allele frequencies and gene diversities were calculated thru gene counting. Allele frequency distributions of the 23 Y STR markers for the Philippine population allele were shown in Figure 3 while summary of gene diversities was presented on Figure 2.

For the Philippine population, gene diversity values for the additional markers (markers that do not belong to the Minimal Haplotype) were found to be on the same range as that of the core haplotype markers (0.441 ± 0.015 to 0.943 ± 0.003) except DYS438 which has the lowest gene diversity (0.253 ± 0.016). Across all markers, DYS385 A/B, a bilocal marker has the highest gene diversity yet among the unilocal markers DYS570 ranks the first (0.853 ± 0.003).



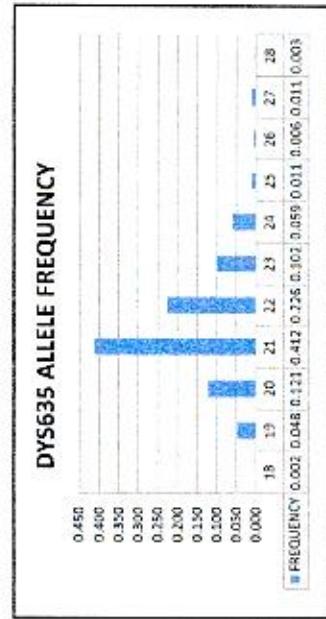
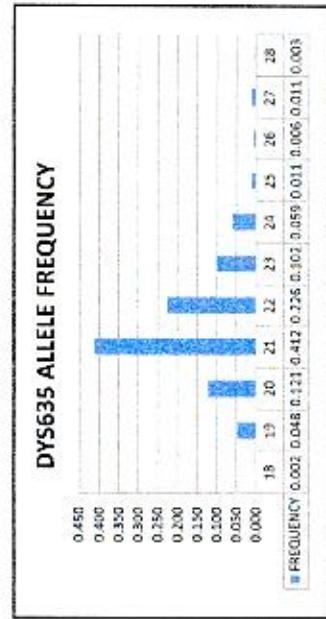
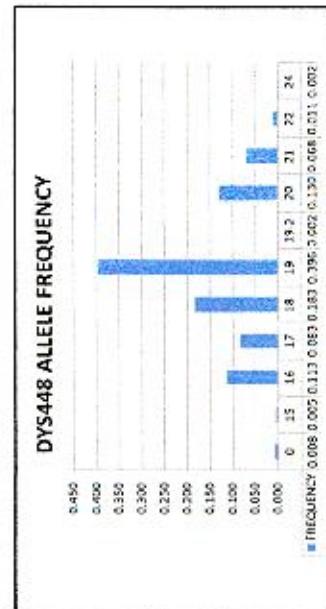
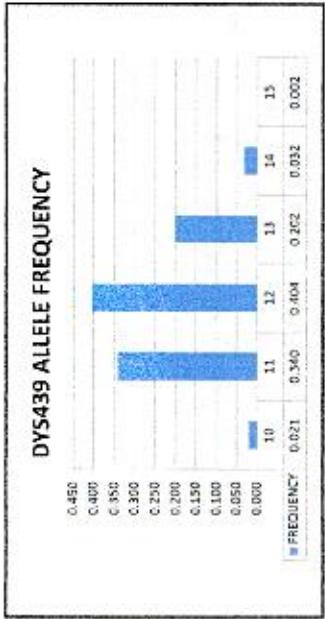
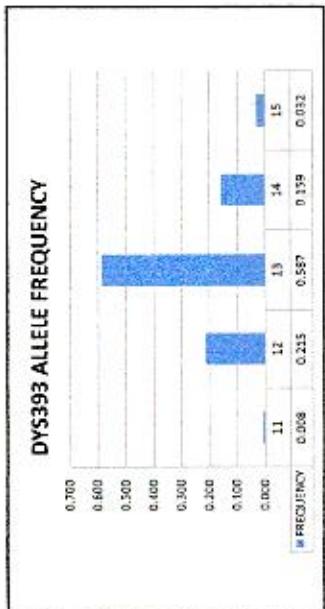
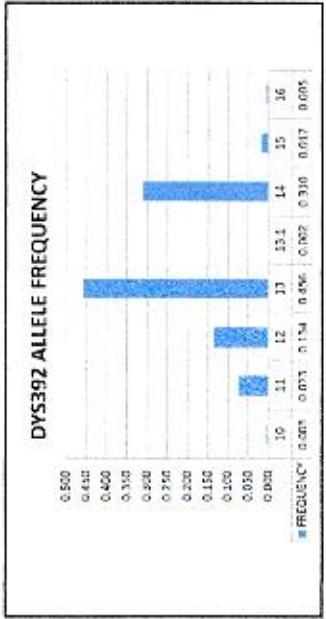
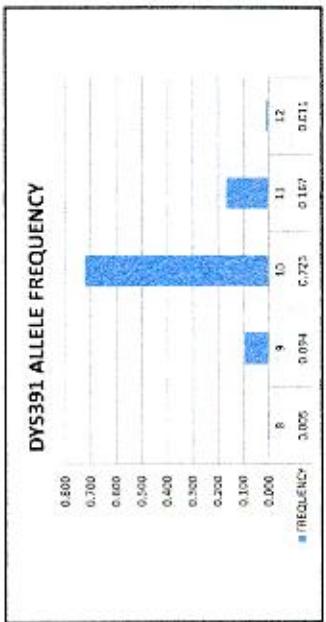
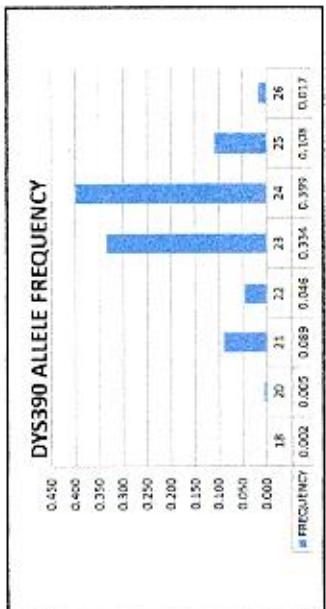


Figure 2. Locus Specific Allele Distribution of 23 Y STR Loci in the Philippine Population

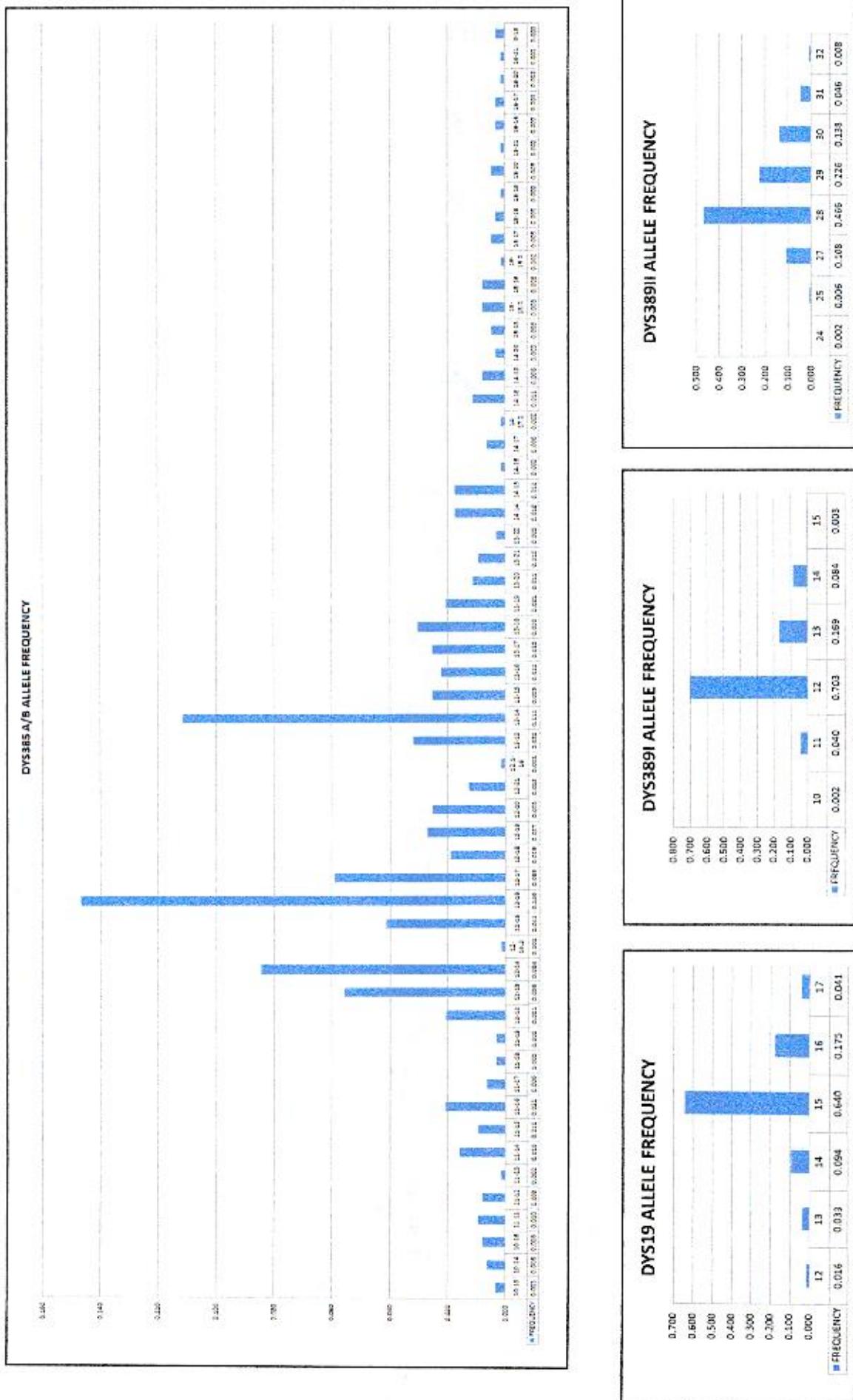
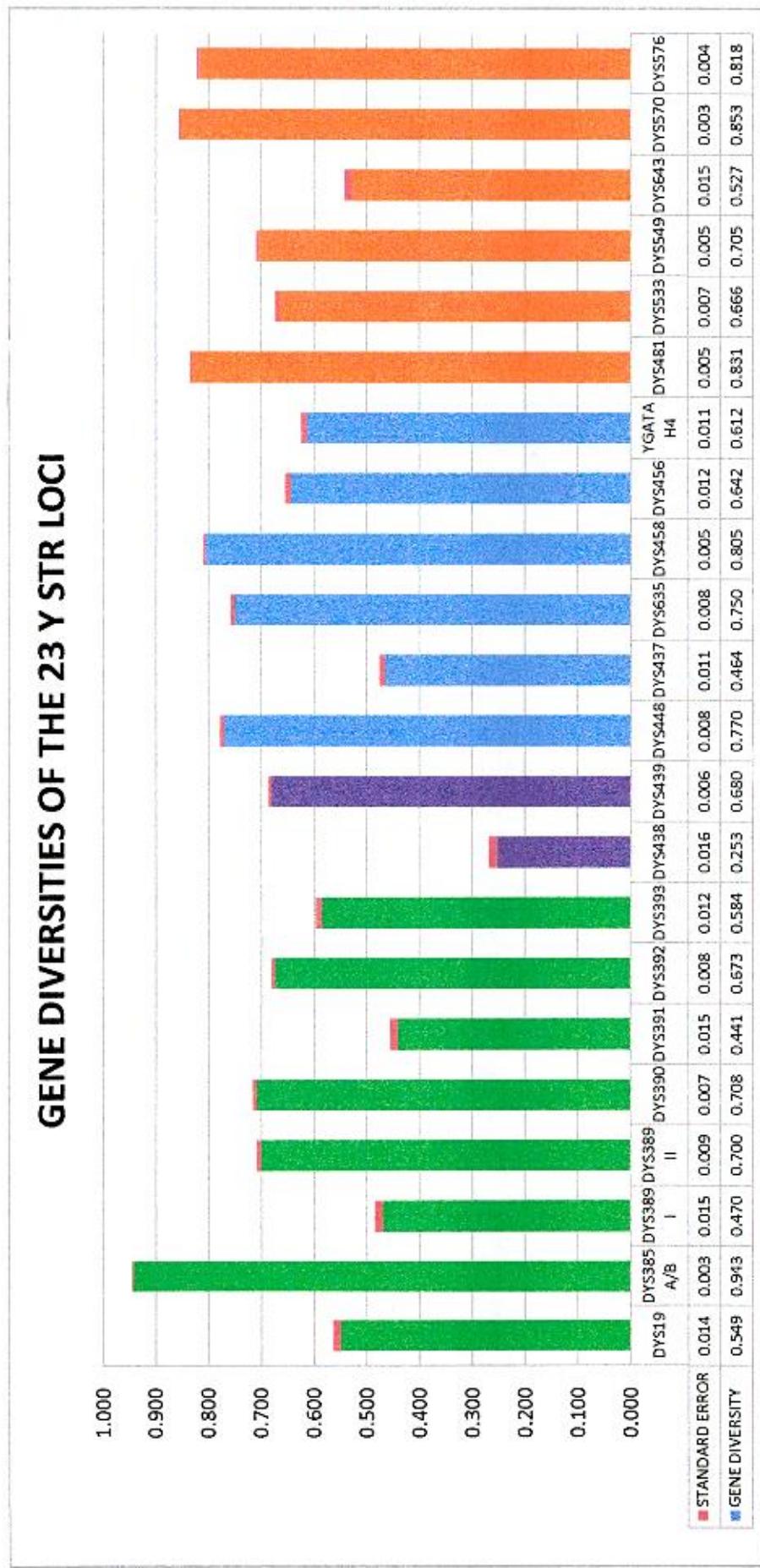


Figure 3. Gene Diversities of the 23 Y STR loci in the Philippine Population



Green = Markers that belong to the European Minimal Haplotype

Green + Violet = Markers forming the Extended Minimal Haplotype

Green + Violet + Blue = Markers included in the Applied Biosystems Y Filer Kit

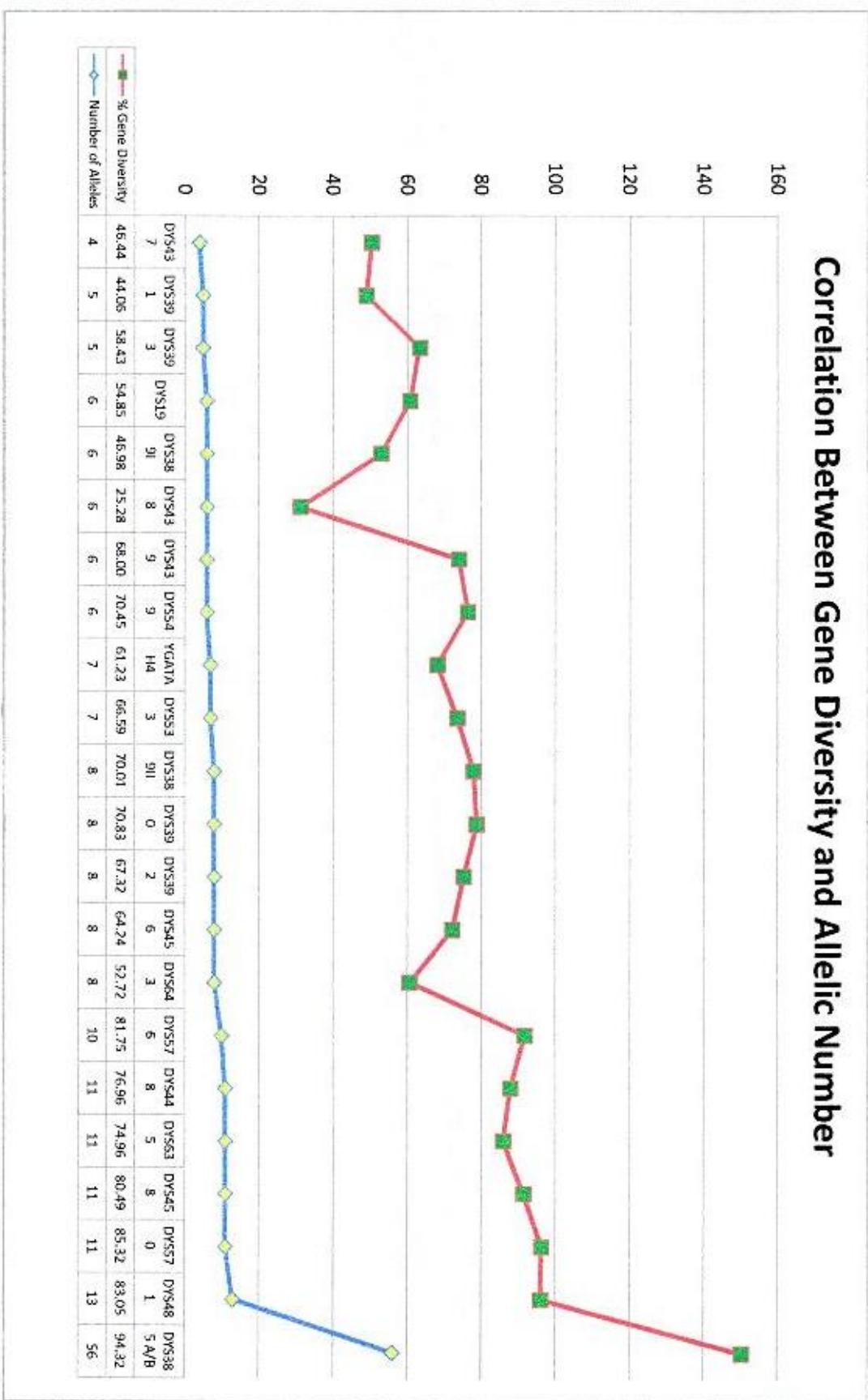
Green + Violet + Blue + Orange= Markers included in the Powerplex Y23 by Promega

Correlation Between Allele Number and Gene Diversities

In order to test if there is a significant relationship between the number of alleles and the gene diversities observed across all loci, Pearson Product Moment Coefficient of Correlation, r was calculated. The obtained r value equals 0.54631 ($r_{\text{calc}} > r_{\text{0.4227}}$) suggest the significant relationship between the allele number and gene diversity. Further, it implies that higher number of alleles renders higher gene diversities Figure 4.

However, this trend does not hold for DYS438 who possess the lowest number of gene diversity equal to 25.28% despite having 6 alleles and to DYS643 possess 52.72% despite having eight alleles. Moreover, DYS576, despite having only 10 alleles has a value higher than the average gene diversity of those markers having 11 alleles. Furthermore, DYS570 having 11 alleles exceeds the gene diversity of DYS481 which has 13 alleles.

Figure 4. Line graph showing the relationship between number of alleles and Gene Diversity. Red line shows the gene diversity trend line while blue line shows the number of alleles trend line



Gene Diversity

Gene diversities between the two data sets have varying measures. DYS19, DYS389II and DYS390 have significantly higher gene diversity for the NCR dataset than that of Philippine population dataset while DYS391 have significantly higher gene diversity than NCR dataset at $\alpha=0.05$. For the unilocal markers, such observation can be attributed to the change in the proportion of the alleles as reflected in Figure 2. Among the loci examined, DYS19 exhibited the greatest decrease ($\Delta GD = 0.091$) in gene diversity in congruence with the increase in the proportion of allele 15. The bialocal marker DYS385A/B do not demonstrate change in the gene diversity despite the complete change in the allele frequency distribution. In addition, it can be noted that no significant difference is observed on the GD between the two dataset for DYS389I, DYS392 and DYS393 at $\alpha=0.05$ while no significant difference exist in all loci except DYS19 at $\alpha=0.01$.

Although there is a similarity in the gene diversity pattern between the two datasets, having DYS385 A/B to express the highest degree of polymorphism and DYS391 express the least, it was observed that DYS19 and DYS393 shifted its positions when ranked according to the level of genetic polymorphism.

Comparison of the Philippine Population Database with NCR Y STR Population Database

Population parameters with respect to the Minimal Haplotype markers (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 AND DYS 385 A/B gathered from this study were compared to the data obtained from pooled Filipino NCR Y-STR database ($n=211$) previously reported by Tan et al (2003) (Table 1).

Table 1. Gene Diversity Values of the Y-STR markers of the Philippine Y-STR Reference Database and the Pooled Filipino NCR Y-STR Database.

Extended Minimal Haplotype Markers	Pooled Filipino NCR Y-STR Database*			Philippine Y-STR DATABASE			Variance	
	Degree of Polymorphism**	GENE DIVERSITY (GD)	Number of Observed Alleles	Degree of Polymorphism	GENE DIVERSITY (GD)	Number of Observed Alleles	$\alpha = 0.05$	$\alpha = 0.01$
DYS19	5	0.640 ± 0.020	6	6	0.549 ± 0.014	6	>	>
DYS385 A/B	1	0.951 ± 0.004	46	1	0.943 ± 0.003	56	≥	≥
DYS389I	7	0.480 ± 0.024	6	7	0.470 ± 0.015	6	≥	≥
DYS389II	3	0.733 ± 0.014	7	3	0.700 ± 0.009	8	>	≥
DYS390	2	0.738 ± 0.012	7	2	0.708 ± 0.007	8	>	≥
DYS391	8	0.380 ± 0.026	4	8	0.440 ± 0.015	5	<	≤
DYS392	4	0.681 ± 0.012	6	4	0.673 ± 0.008	8	≥	≥
DYS393	6	0.597 ± 0.018	4	5	0.584 ± 0.116	5	≥	≥

**Markers were ranked according to the level of genetic polymorphism based on GD; rank 1 exhibits the highest degree while rank 8 exhibits the least

*Values obtained from Tan et al, 2003

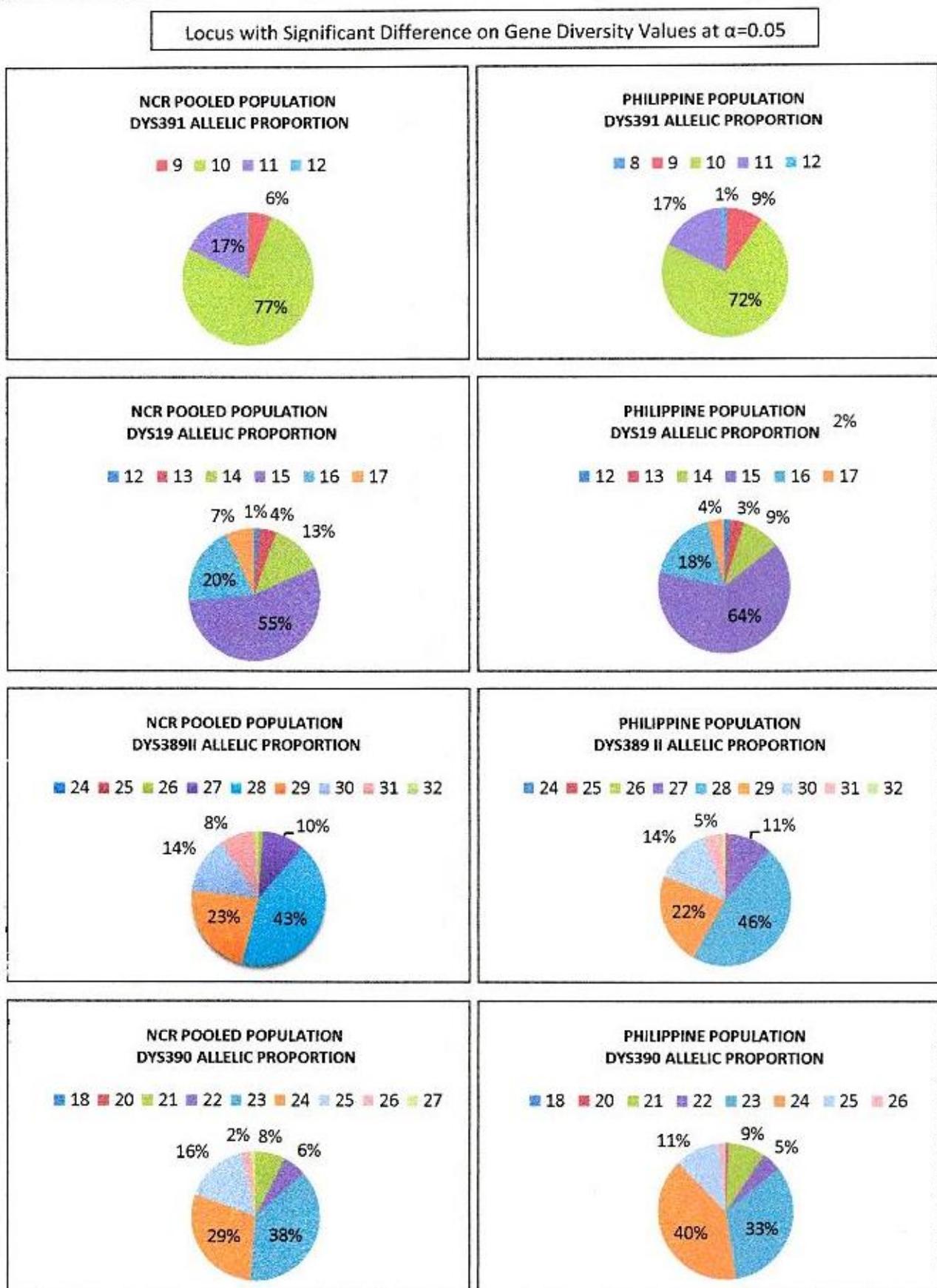
> Gene diversity of Pooled Filipino NCR Y-STR Database is significantly greater than the Philippine Y STR Database

≥ Gene diversity of Pooled Filipino NCR Y-STR Database is greater but not significantly different to Philippine Y STR Database

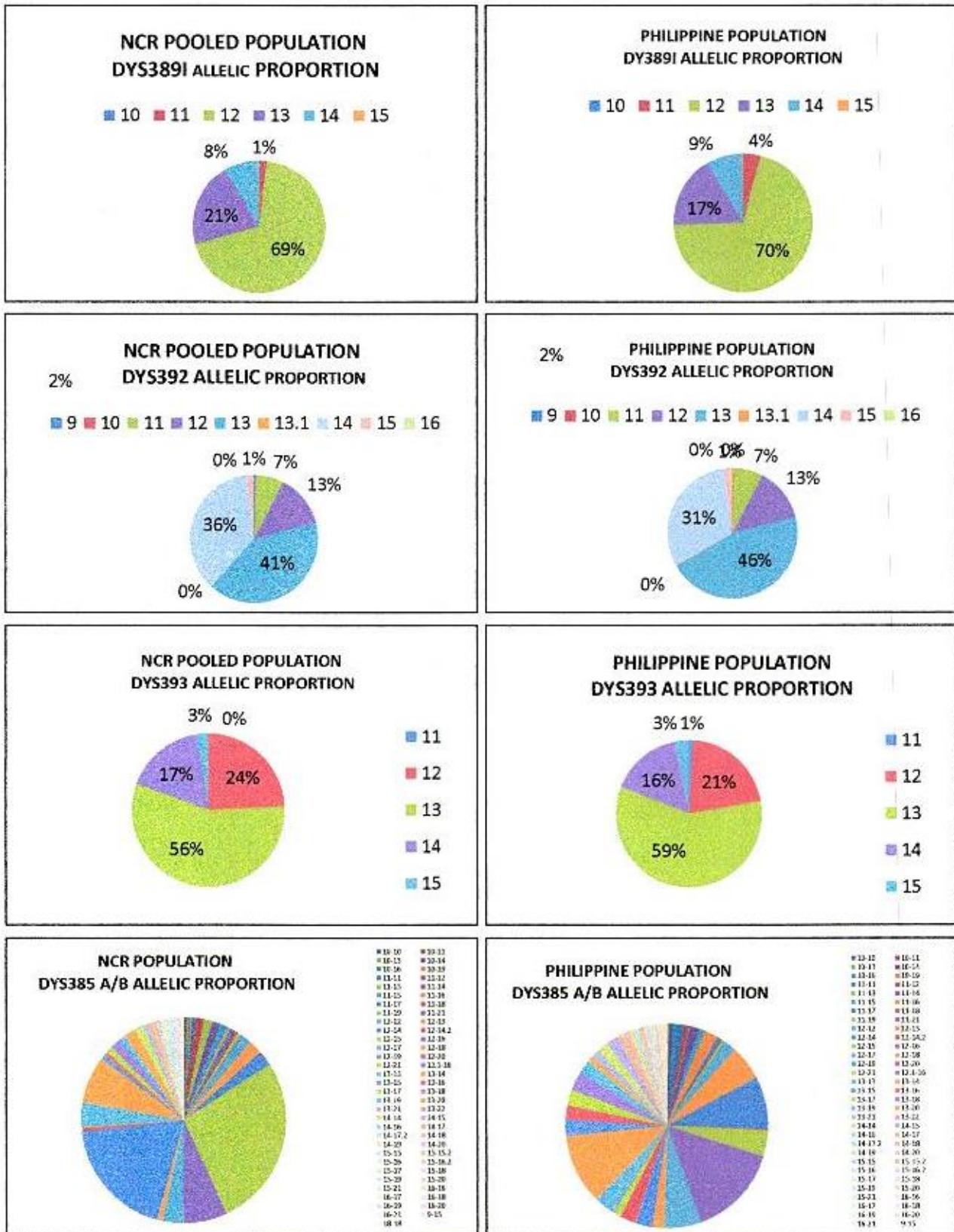
< Gene diversity of Pooled Filipino NCR Y-STR Database is significantly greater than the Philippine Y STR Database

≤ Gene diversity of Pooled Filipino NCR Y-STR Database is greater but not significantly different to Philippine Y STR Database

Figure 5 A. Comparison of Allelic Proportion Between Pooled NCR Population and Philippine Population



Locus with No Significant Difference on Gene Diversity Values at $\alpha=0.05$



Rare Alleles

Comparing the number of observed alleles, it can be noted that more alleles were observed in the Philippine population dataset than the NCR dataset as expected due to increase in population size as well as increase of the sampling area (Table 2). New allele combinations were observed at DYS 385 A/B. The list of the new alleles as well as rare alleles (defined as alleles with frequency < 0.01 (Kimura, 2002)) detected on the newly Philippine Population Dataset are shown in Table 3.

Table 3. List of the newly detected rare alleles on the Philippine population.

LOCUS	New/ Rare Allele (frequency)		LOCUS	New/ Rare Allele (frequency)	
DYS389 II	24 (0.002)	25 (0.006)	Y GATA H4	15 (0.002)	9(0.002)
DYS390	18 (0.002)	20 (0.005)	DYS533	15 (0.002)	9 (0.005)
DYS391	8 (0.005)		DYS549	15 (0.003)	
DYS392	10 (0.003)	13.1 (0.002)	DYS458	15.1 (0.002)	21 (0.005)
	16 (0.005)			22 (0.002)	
DYS393	11 (0.008)		DYS643		8 (0.005)
DYS385A/B	12,14.2(0.002)	12.1, 16 (0.002)	DYS456		17 (0.003)
	13,21* (0.01)	13,22 (0.003)			19 (0.003)
	14,17.2 (0.006)	14,19 (0.008)			20 (0.002)
	15,15.2 (0.008)	15,16 (0.008)			30 (0.002)
	15,16.2 (0.002)	15,19 (0.002)			
	15,20 (0.005)	15,21(0.002)			
	16,17 (0.003)	16,20 (0.002)			
	16,21 (0.002)	9,15 (0.003)			
DYS438	13 (0.002)		DYS570	23 (0.005)	
DYS439	15 (0.002)		DYS576	13 (0.002)	22 (0.002)
DYS448	15(0.005)	19.2 (0.002)	DYS456	13 (0.008)	19 (0.006)
	24(0.002)			9* (0.002)	
DYS437	3 (0.005)				
DYS635	18 (0.002)	26 (0.006)			
	28(0.003)				

*13,21 Allele in the DYS385 A/B was not detected on the NCR pooled database

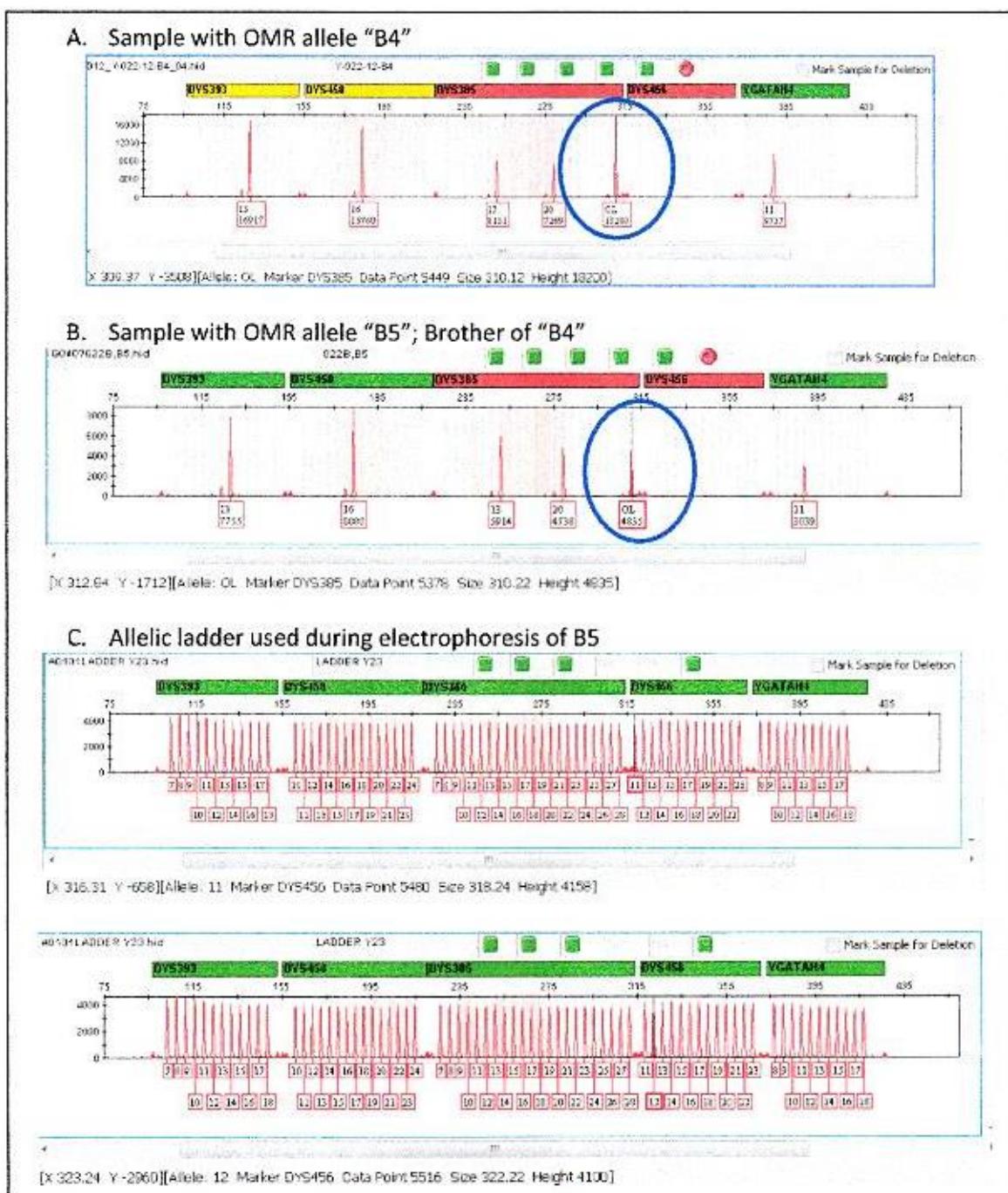
** At Off Marker Range was detected using the PPY23; allele sequencing resulted to an allele designation = 9

Variant Alleles

An off marker range (OMR) allele was detected at the DYS456 locus during the typing process.

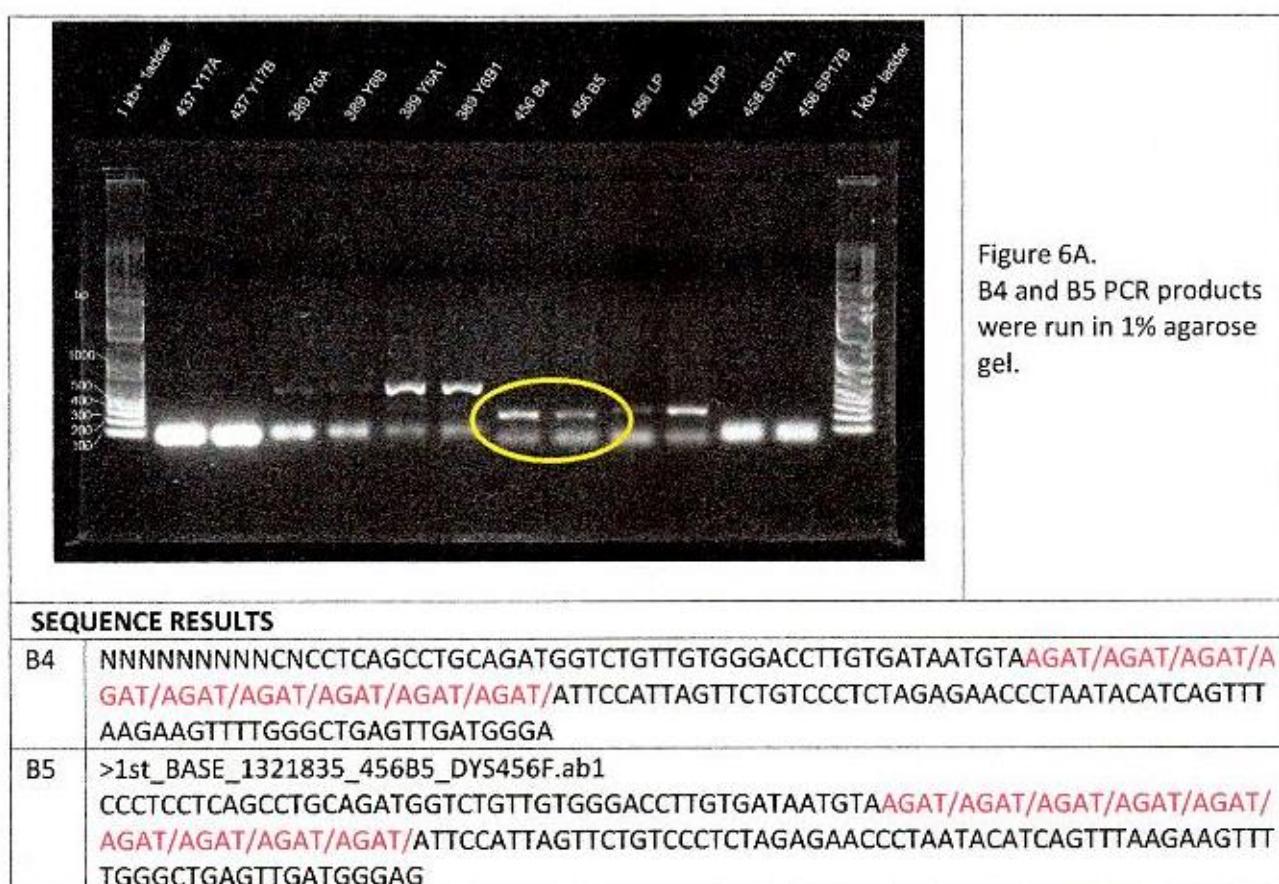
The same sample was amplified and typed using the Y Filer system which yielded the same result. During the review of the sample, we found out that DNA sample of his brother is also available, thus both samples were analyzed. Results of the runs are displayed on Figure 6.

Figure 6. Electropherogram of Samples Exhibiting Off Marker Range Allele



Allele designation of OMR alleles was first estimated based on the allelic size. The shortest allele (allele 11) on the ladder of DYS456 has a size of 318.24 while the second to the shortest allele (allele 12) has a size of 322.22. Getting its difference gives the approximate size of a single repeat ($\Delta = 3.98$). Subtracting this value to the last allele gives the approximate size of the next shorter allele (allele 10) and subtracting twice the size of a single repeat yields the approximate size of the OMR allele (310.22) and the number of repeat is designated to be nine. Same method was employed on "B4" and similar result was obtained.

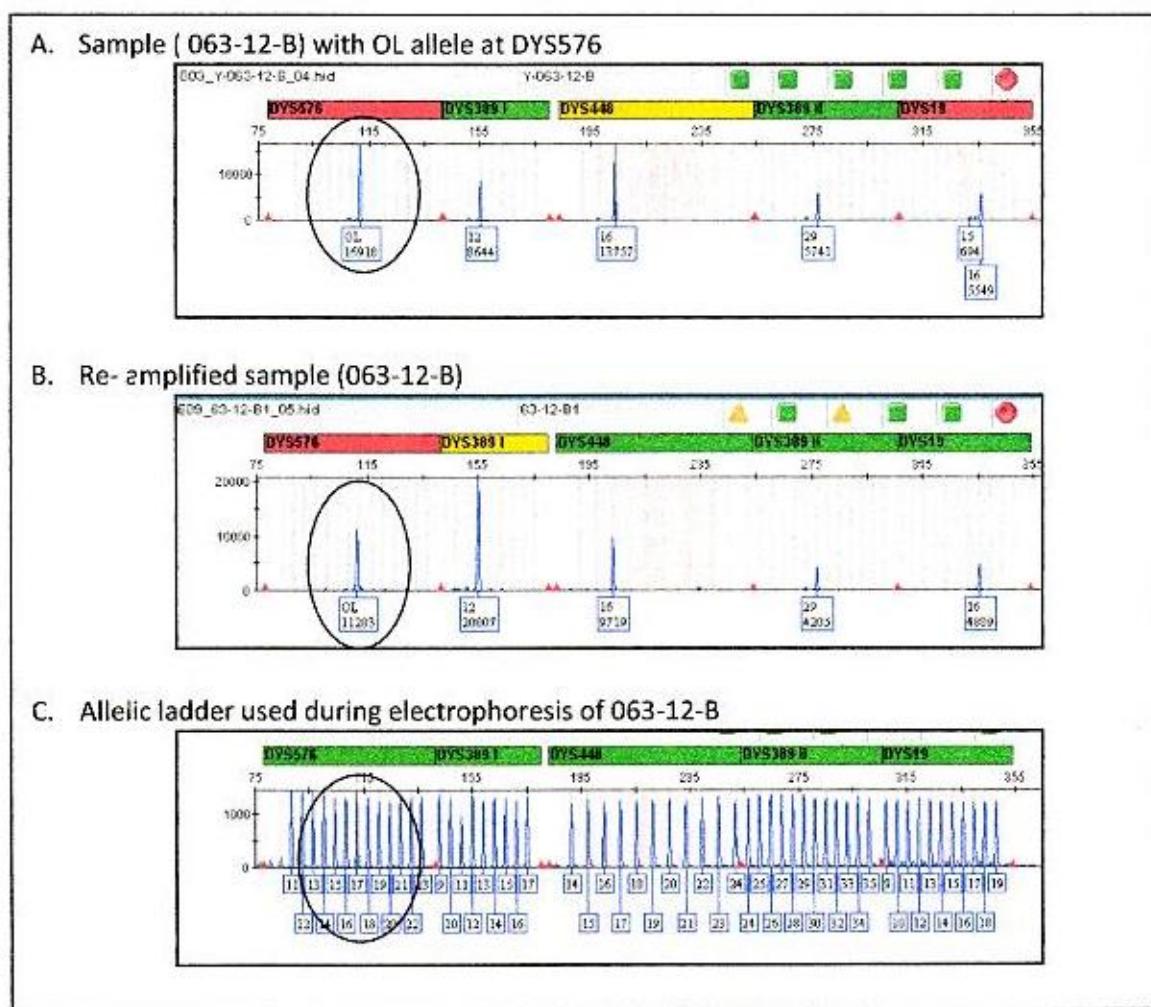
Allele designation was further confirmed by allele sequencing. Both samples were reamplified using non labelled primer (primer sequence found on Appendix C (Hill, 2008)) following the procedure described in the methodology. Presence of amplified product was tested by running the PCR product in a 1% agarose gel (Figure 6A). Sequencing confirmed that both B4 and B5 contain nine repeats.



In addition to the variant allele, a microvariant and was detected in sample 063-12-A. A microvariant is said to be a variation of the allele that contains incomplete repeat. During typing, an off ladder allele peak with a size equal to 111.25 was detected on the said sample on DYS576. The sample was reloaded and reamplified, yet same result displayed on Figure 7 was obtained.

Allele designation of the off ladder allele was again estimated based on the allelic size. The off ladder peak was found to be in between peaks 16 and 17 (as compared with the allelic ladder) with a corresponding size of 108.37 and 112.40. Getting its difference gives the approximate size of a single repeat ($\Delta = 4.03$). Since this marker is a tetranucleotide repeat, each base has a corresponding size of 1.0075. Subtracting this value to the size of allele 17 yields the size of the approximate size of the allele in question which suggest that our allele is 1 base shorter from allele 17, thus the OL peak was designated as 16.3.

Figure 7. Electropherogram of Samples Exhibiting Off Marker Range Allele



Additional Alleles

In YSTR analysis, presence of additional peaks in markers other than DYS385 A/B implies the presence of male mixtures. However, instances of duplicated, triplicated or even quadruplicated alleles have been reported (Decker et al, 2008; Diederiche et al, 2005; Ballard et al, 2005; YHRD, <http://www.yhrd.com>). In addition, Diederiche et al (2005) have described a sample with duplications at three different loci. Occurrence of such complicates evidence interpretation. Thus, in the course of forensic evaluation, DNA analysts must be aware that presence of duplicate alleles may not suggest mixture. Furthermore, accurate data on the presence of these alleles as well as identification of markers exhibiting this features must be gathered and published to prevent misinterpretation.

During father and son transmission studies, we have observed a duplication at DYS570 wherein the father and his son have a score of 19,20. In order to confirm, sample was re-extracted, reamplified and reanalyzed yet identical result was obtained. Further confirmation was done by reprocessing the sample using PP21 to confirm that none of the samples were contaminated. The YHRD database (YHRD, <http://www.yhrd.com>) reflects only one record of duplication at this locus and no record of having this score. Electropherogram for both samples are reflected on Figure 8.

Meanwhile, a quadruplicate allele was detected at DYS385 A/B. Same with the previous sample, contamination was excluded by re-extraction, re-amplification and re-analysis using both Y STR and autosomal analysis. No extra peaks were observed on both process. In addition, the sample was analyzed using the Y filer kit and yielded the same result. No record of having this score was reported at the YHRD database yet.

Figure 7. Electropherogram of Samples Exhibiting Duplicate allele

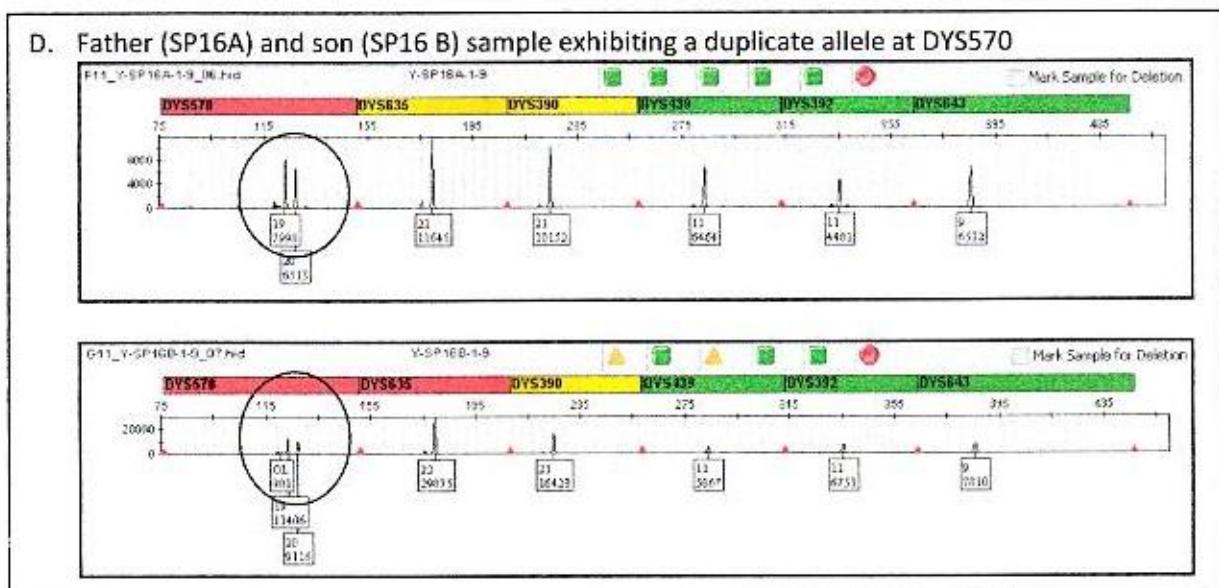
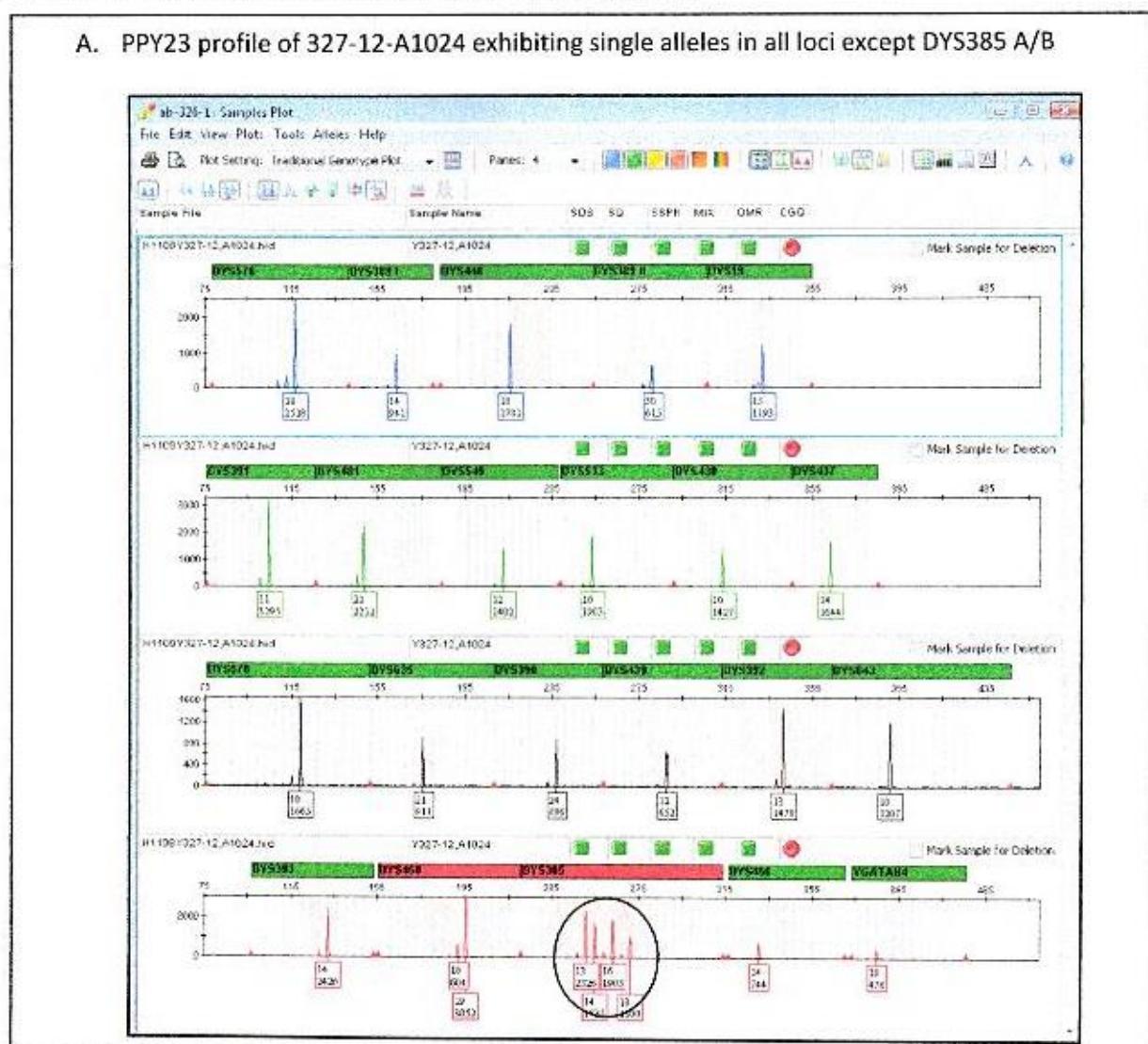


Figure 7. Electropherogram of Samples Exhibiting Quadruplicate allele



Haplotype Analysis involving Minimal Haplotype Loci

In order to compare our data with the NCR pooled database (N=211) (Tan et al, 2003), the Philippine population data set (N=629) was examined using the European minimal haplotype marker set. A total of 434 haplotypes out of the 629 individuals were discriminated after the process. This entails the presence of 347 unique haplotypes, 54 haplotypes shared by 2 individuals, 17 haplotypes shared by 3 individuals, 5 haplotypes shared by 4 individuals, 2 haplotypes shared by 5 individuals, 2 haplotypes shared by 6 individuals, 1 haplotype is shared by 7 individuals, 2 haplotypes shared by 8 individuals and 1 haplotype is being shared by 9,13,14 and 23 individuals respectively.

The haplotype diversity obtained has a value of 0.99637 with a standard error of 0.00043 and was found to be not significantly different from the values obtained by Tan et al. Discrimination capacity was found to be 70.00%.

Forensic Assessment Involving the Comparison of Four Haplotype Sets

One of the contentions in Y STR typing involves the number of Y STR loci to be typed. It has been mentioned by several authors (Belleza et al, 2003 and Gusmao et al, 2002) that the non recombining nature of Y chromosomes renders the inability of additional Y STR markers despite its inherent superior polymorphic property to discriminate profiles due to linkage. Thus, it is imperative that we assess the consequence of increasing markers in discriminating the profiles. In order to do this, haplotype diversity and discrimination capacity for each Y STR sets (Minimal haplotype, Extended Minimal Haplotype, Y Filer and PPY23) were calculated and displayed on Figure 8 .

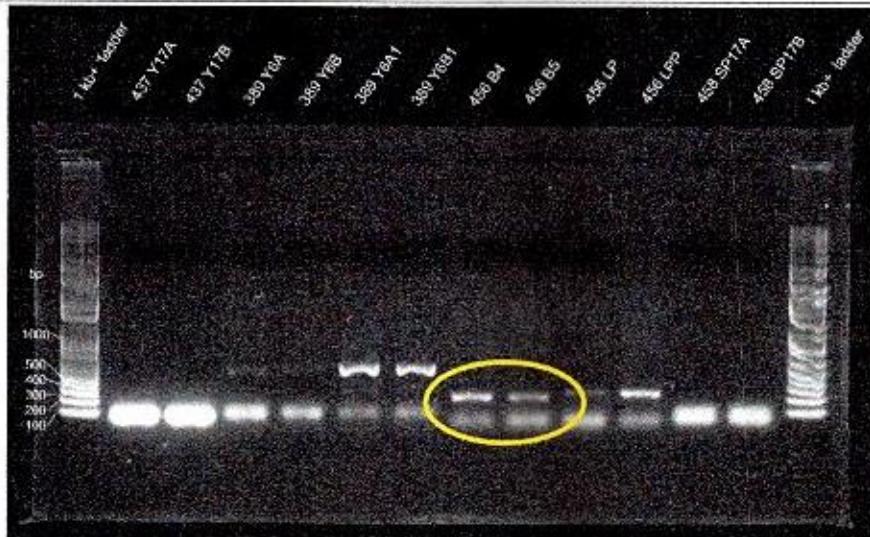


Figure 6A.
B4 and B5 PCR products
were run in 1% agarose
gel.

SEQUENCE RESULTS

B4	NNNNNNNNNCCTCAGCCTGCAGATGGTCTGTTGGGACCTGTGATAATGTAAGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/ATTCCATTAGTTCTGTCCTCTAGAGAACCTAATACATCAGTTAAGAAGTTTGGGCTGAGTTGATGGGA
B5	>1st_BASE_1321835_456B5_DYS456F.ab1 CCCTCCTCAGCCTGCAGATGGTCTGTTGGGACCTGTGATAATGTAAGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/AGAT/ATTCCATTAGTTCTGTCCTCTAGAGAACCTAATACATCAGTTAAGAAGTTTGGGCTGAGTTGATGGGAG

Haplotype Analysis involving 23 Y STR Loci

Haplotype analysis of the Philippine population involving 23 Y STR markers yields a haplotype diversity of 0.99992 with a standard error of 0.00230. This defines the presence of 612 haplotypes from the sample of 629 individuals. Out of the 612 haplotypes, 596 haplotypes were found to be unique, 15 haplotypes were found to be shared by two individuals and 1 haplotype was shared by three individuals. Haplotype list is reflected on Appendix B while frequent haplotypes is shown in table Based on this data, the discrimination capacity was calculated to be 97.30%.

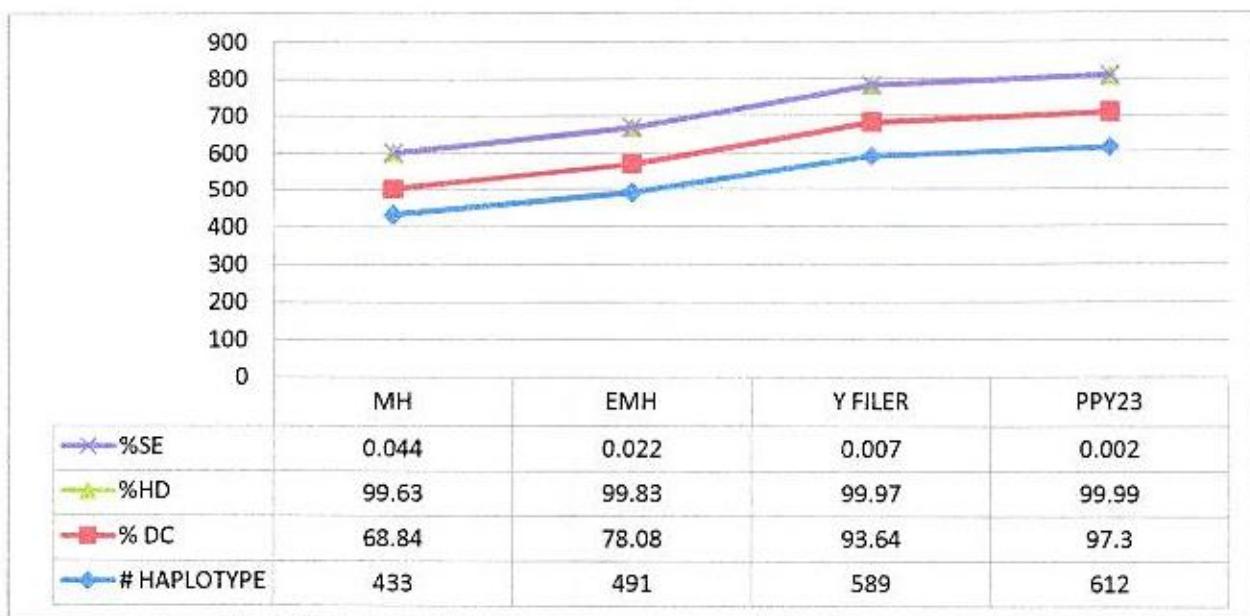
Null Alleles

Ten samples produced null alleles during the typing process as summarized in Table 4. Samples were reprocessed using PPY23 to remove the possibility of a mere allele drop out. Primer binding site mutation on markers DYS448, DYS19 and DYS437 were excluded by using Y filer system. Five samples that failed to amplify at DYS448 using the PPY23 produced the same result when typed with Y Filer which suggest the absence of primer binding site mutation. On the other hand, a sample that initially failed to amplify at DYS19 using the PPY23 produced a product (allele = 15) using the Y filer kit therefore suggesting mutation at the primer binding position.

Table 4. Summary of Observed null alleles

SAMPLE CODE	PLACE OF ORIGIN	FOCUS	Confirmatory Test
Y-CIC025-11	Taguig City, NCR, Philippines	DYS576	
230C-11-N	Albay/Region 5, Philippines	DYS448	Y filer/ null allele
012-12-A	Pangasinan/Region 1, Philippines		Y filer/ null allele
300-12-A859	Pangasinan/Region 1, Philippines		Y filer/ null allele
332-12-A1066	Ilocos Norte, Region 1, Philippines		Y filer/ null allele
MAV-79	Manila / NCR, Philippines		Y filer/ null allele
PNPDB-318	Quezon City / NCR, Philippines	DYS19	Y filer/ 15
Y-CORBC-003-13B008	Baguio/Cordillera Region, Philippines	DYS481	
Y-CIC-040-11	Cebu, Region 7, Philippines	DYS437	
Y-CICDNA-007-11	Zamboanga, Region 9, Philippines	DYS643	

Figure 8 . Forensic Assessment of 4 Haplotype Sets



From the chart, we can see that increasing loci results to an increase in the discrimination capacity as well as haplotype diversity. This shows that for the Philippine population, having a multiplex set with more Y STR markers takes an advantage in individualizing a profile than using a Y STR set with lower number of markers. This clearly manifests the forensic usefulness of the 12 additional markers taken together in discriminating the haplotypes rendering a significant increase on the haplotype diversity at $\alpha=0.05$ from 99.63% to 99.99%. In addition, 178 additional haplotypes were successfully resolved with the addition of 12 markers from the extended minimal haplotype markers.

However, despite the increase in haplotype diversity and discrimination power, we can see that a multiplex system with 23 markers still is not able to individualize the profiles having only 612 haplotypes out of the 629 individuals. Unresolved haplotypes are listed in Table 5. This implies that additional or more improved markers are required. Furthermore, comparison of

the four sets exemplifies the advantage of extending the YSTR database set in order to come up with a more relevant data for probabilistic calculations.

Table 5. List of the Common Haplotypes from the Philippine Y STR Population Database

HAPLOTYPE NUMBER	DYS576	DYS289I	DYS448	DYS389II	DYS19	DYS301	DYS441	DYS549	DYS533	DYS438	DYS337	DYS570	DYS535	DYS390	DYS539	DYS392	DYS443	DYS393	DYS458	DYS385 A/B	DYS456	YCA7AH4	N	FREQUENCY
462	19	11	19	27	15	10	24	11	13	10	15	14	21	24	12	13	11	13	16	12,16	15	12	3	0.00477
27	15	12	17	29	14	9	25	13	11	10	14	18	22	21	11	12	11	13	14	12,14	15	13	2	0.00318
99	16	12	17	28	16	10	25	12	11	10	14	19	20	24	12	12	11	14	14	12,14	15	11	2	0.00318
117	16	12	18	28	15	10	24	12	11	10	14	19	19	23	11	14	12	13	16	12,13	17	11	2	0.00318
190	17	12	18	27	14	10	23	13	11	11	15	19	28	23	12	14	11	13	16	12,14	15	13	2	0.00318
230	17	12	19	28	15	11	25	11	13	10	14	16	22	24	13	13	11	12	16	12,13	15	12	2	0.00318
252	17	12	21	30	16	11	25	12	11	10	14	19	21	24	11	13	11	13	19	12,19	14	12	2	0.00318
273	17	13	19	29	15	11	24	11	13	10	14	15	22	24	13	13	11	13	16	12,15	15	12	2	0.00318
287	17	14	20	31	12	10	24	12	11	10	15	15	21	24	12	13	12	13	17	13,17	14	11	2	0.00318
295	18	11	19	27	15	10	24	11	13	10	15	15	21	24	12	13	11	13	16	12,16	15	12	2	0.00318
477	19	12	19	28	15	10	21	11	12	10	14	15	22	24	13	13	11	13	17	13,17	15	13	2	0.00318
490	19	12	19	28	15	10	24	11	12	10	15	14	25	24	13	13	10	13	16	12,16	15	12	2	0.00318
535	19	13	18	29	15	10	24	12	11	10	15	18	21	22	11	13	10	12	21	11,11	15	12	2	0.00318
567	20	12	19	28	15	10	24	11	13	10	15	14	21	24	13	13	11	13	18	12,16	15	13	2	0.00318
574	20	12	20	28	14	10	23	12	11	11	15	18	21	24	13	14	12	12	18	12,17	15	12	2	0.00318
594	20	14	17	31	15	11	22	13	10	10	14	17	21	25	11	13	12	15	17	13,19	16	10	2	0.00318

Forensic Evaluation of Six additional Y STR Markers

To determine whether each of the 6 extra loci of the 23 Y STR system will increase the capacity of the 17 Y STR system set to discriminate more profiles, each marker was added to the 17 Y marker set. Data on the haplotype diversity, discrimination capacity and number of resolved haplotypes were gathered and evaluated. The graph (Figure 9) showing increase in the haplotype diversity, discrimination capacity and # of resolved haplotypes when compared with the 17 Y STR system set alone demonstrates the value of adding any of the markers. Also, the graph shows that DYS570 and DYS576 is most useful when added.

Figure 9. Forensic evaluation of six additional markers

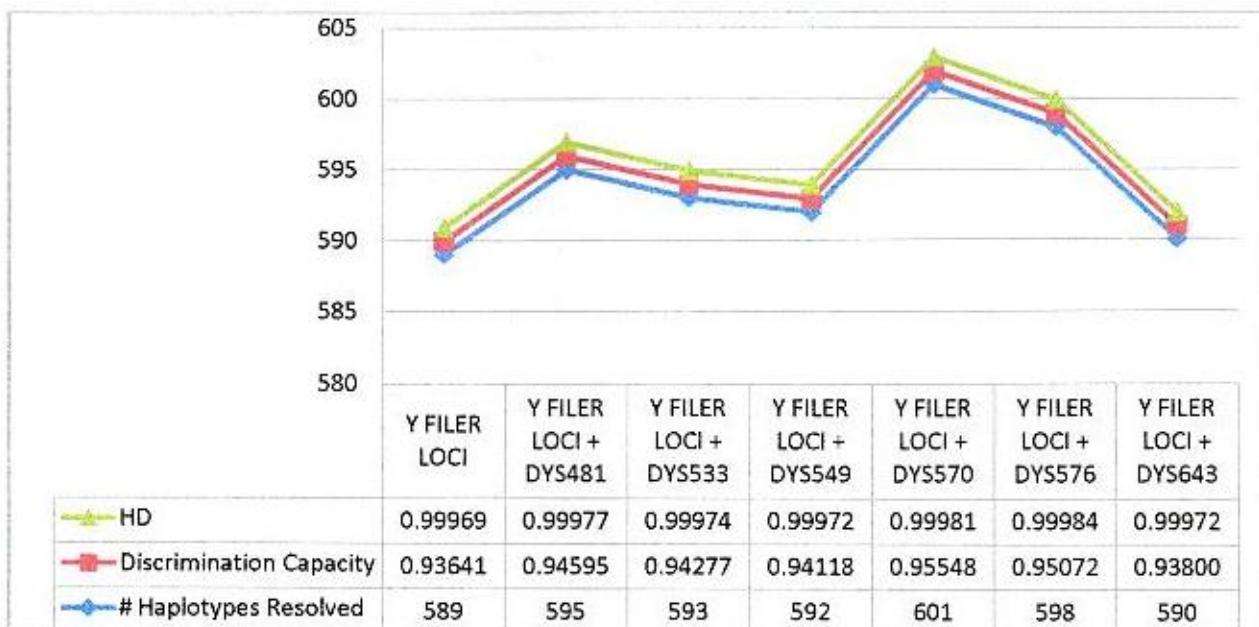
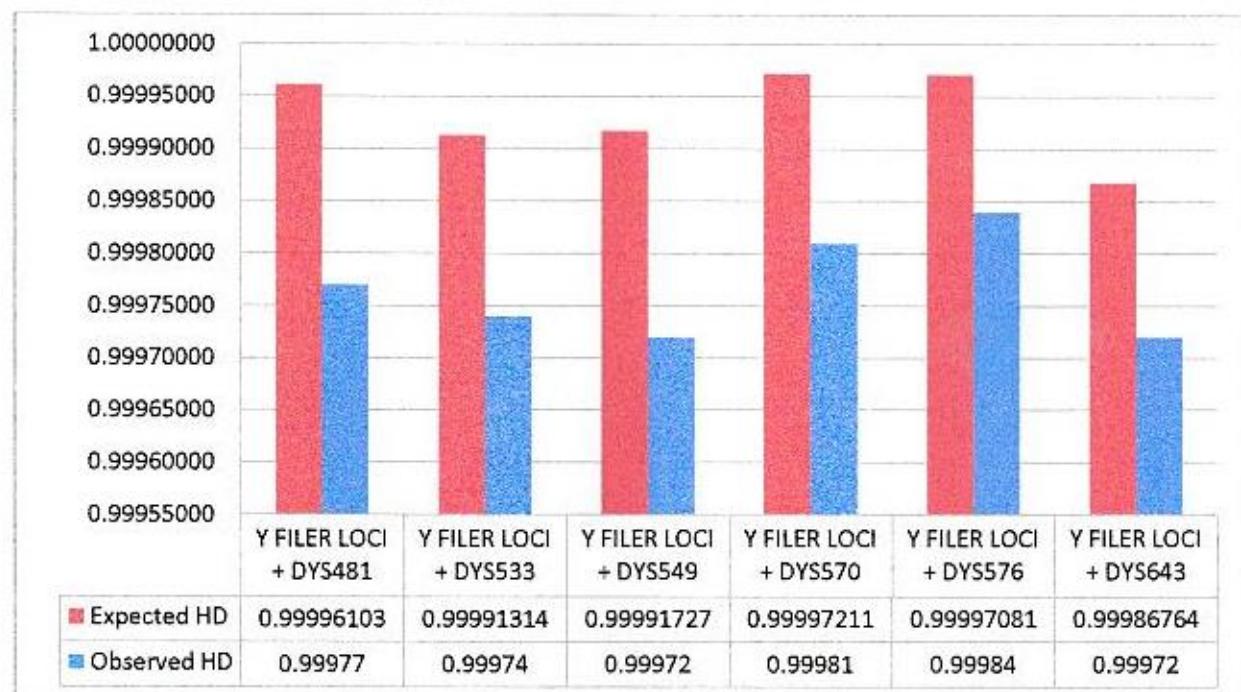


Figure 9A. Observed and Expected haplotype diversity with the addition of extra Y markers of PPY23 to the Yfiler set.

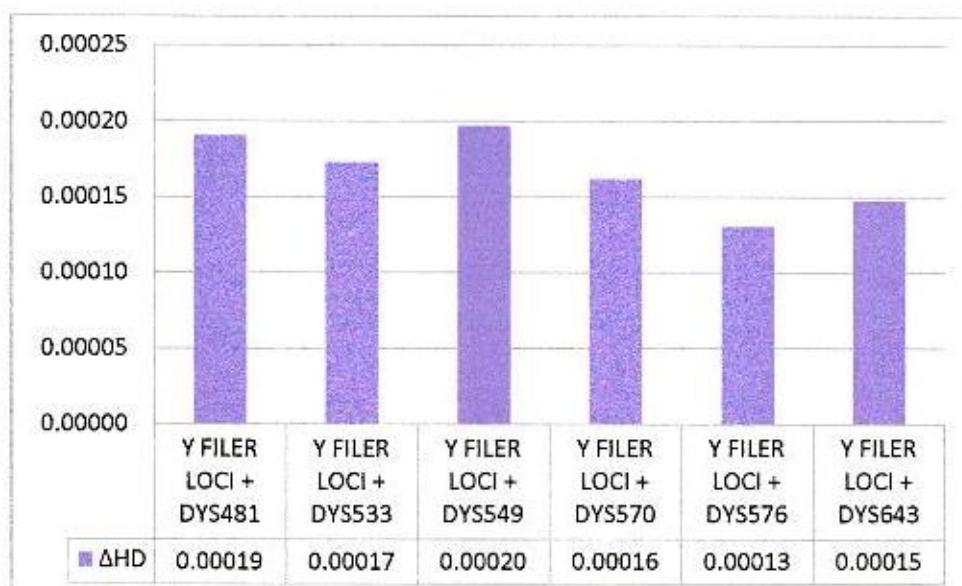


In order to compare the individual locus informativeness of extra Y markers incorporated in the PPY23 multiplex system versus the effective contribution of this markers to

the Y filer set, Expected Haplotype Diversity was measured. Expected haplotype Diversity was defined to be the “theoretical value calculated as if the locus under analysis were totally unlinked” (Gusmao et al, 2002) to the haplotype set and was calculated using the CPD formula.

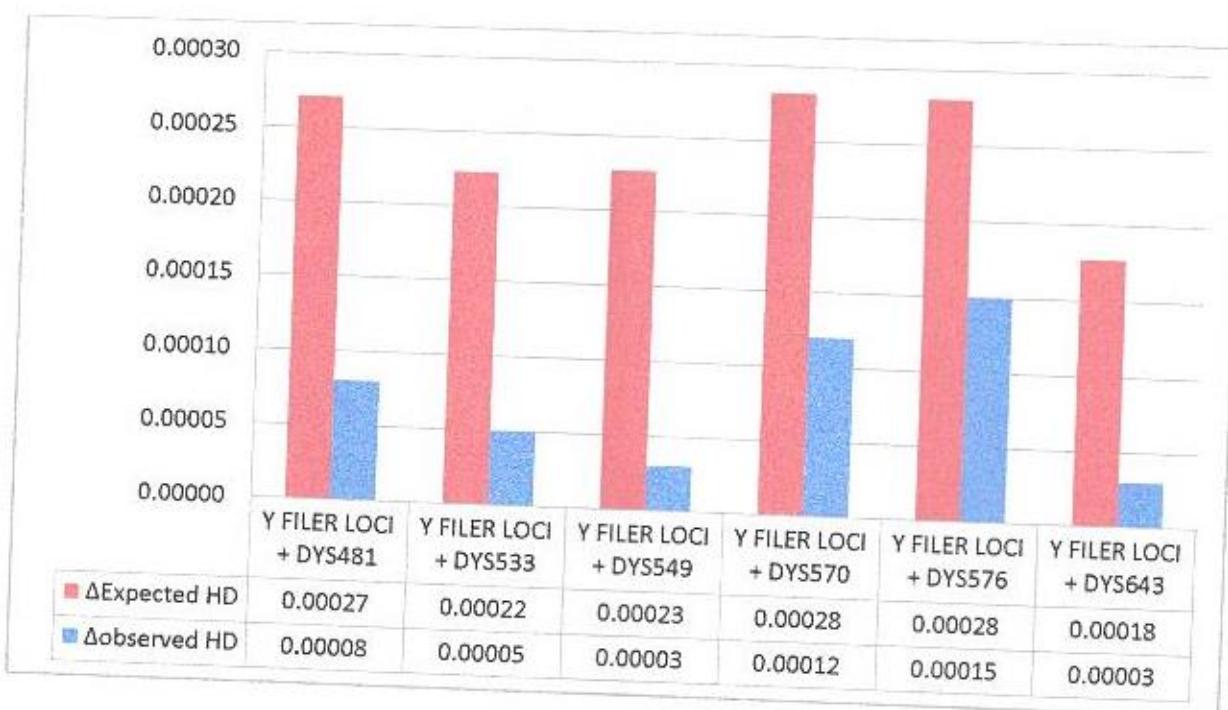
A lower value for the observed haplotype diversity when compared with the expected haplotype diversity which reinforces the concept that linkage exist between the markers included in the haplotype with the marker being analyzed (Figure 9A). In addition, subtracting the value of observed HD from the expected HD determines the degree of linkage that occurs between the additional marker with the Y Filer set. From figure 10, we can see that the greatest linkage (highest deviation) occurs at DYS549 while smallest linkage effect (least deviation) is observed at DYS576. This method shows a pragmatic approach in measuring the linkage between marker sets.

Figure 10. Difference Between Expected and Observed Haplotype Diversity with the addition of extra Y markers of PPY23 to the Yfiler set .



Change between the expected haplotype diversity with the Y Filer set haplotype diversity measures the individual informativeness of the locus being analyzed while change between the observed haplotype diversity with the Y Filer set haplotype diversity quantifies the effective contribution of the locus being analyzed. Comparing this two values as shown in figure 11 shows that an inherently informative locus may not be able to contribute much in the haplotype diversity when combined with other marker set.

Figure 11. Observed and Expected Increase in the haplotype Diversity with the addition of extra Y markers of PPY23 to the Yfiler set .



Mutation Analysis

One hundred and ten father and son transmission from previously confirmed paternity related cases were screened for mutation. This has a total of 230 profiles comprising of 97 fathers and 133 sons. Out of these samples, ten mutational events have been recorded and summarized in table 5. A sample

which exhibited mutation in both 389I and 389II were counted as single mutational event considering that DYS389I is a subsection of DYS389II (Decker et al , 2008 and Ge et al, 2009).

A balance of gain (5) and loss (6) of repeat was also observed showing that these Y chromosome microsatellites neither has expansion nor contraction bias similar to the observation of Ge et al (2009). All of the mutational events were found to be single step. Review of the autosomal records show that none of these samples displayed mutation in 21 STR markers.

Locus specific mutational rates as well as overall mutation rates have been calculated by dividing the number of observed mutational event over the number of transmissions. The obtained values were compared with the values recorded in the YHRD database.

The calculated overall mutation rate has a value of 3.6×10^{-3} (95% CI $(1.8-6.4) \times 10^{-3}$). Based on the confidence interval, mutation rate is equal to the mutation rates reported by Decker et al (2008) (3.6×10^{-3}) and similar to values obtained by Ge et al (2009)(3.6×10^{-3}) yet significantly higher to the mutation rate reported by Budowle et al (2005) 1.57×10^{-3} .

Locus by locus comparison of the observed mutation rates with that of the values collected by the YHRD proves to be similar based on the confidence interval. In addition, interlocus comparison showed that there is no significant difference. However, when point estimates of the mutation rate is being considered, DYS635 shows the highest mutation rate 2.26×10^{-2} (95% CI $(4.7-64.5) \times 10^{-3}$) in the Philippine population while DYS570 demonstrates the highest in the YHRD database.

Table 5. Comparison of Mutation count, rates and 95% Confidence interval for the Philippine Population with the YHRD published data.

PHILIPPINE POPULATION DATABASE						YHRD			
Locus	Mutations	Meloses	Mutation Rate ($\times 10^{-3}$)	95% CI ($\times 10^{-3}$)	Locus	Mutations	Meloses	Mutation Rate ($\times 10^{-3}$)	95% CI ($\times 10^{-3}$)
DYS456	1	133	7.519	0.2-41.2	DYS456	28	6,678	4.193	2.78-6.05
DY389I	0	133	0.000	0-27.4	DY389I	37	13,788	2.683	1.89-3.69
DYS390	0	133	0.000	0-27.4	DYS390	31	15061	2.058	1.40-2.92
DYS389II	1	133	7.519	0.2-41.2	DYS389II	52	13759	3.779	2.82-4.95
DYS458	1	133	7.519	0.2-41.2	DYS458	45	6,677	6.740	4.92-9.01
DYS19	0	133	0.000	0-27.4	DYS19	36	15,539	2.317	1.62-3.21
DYS385	1	133	7.519	0.2-41.2	DYS385	59	25,620	2.303	1.75-2.97
DYS393	0	133	0.000	0-27.4	DYS393	13	13,713	1.904	6.12-1.80
DYS391	1	133	7.519	0.2-41.2	DYS391	38	14,935	2.544	1.80-3.49
DYS439	0	133	0.000	0-27.4	DYS439	54	10,096	5.349	4.02-6.97
DYS635	3	133	22.556	4.7-64.5	DYS635	28	7,525	3.721	2.47-5.37
DYS392	0	133	0.000	0-27.4	DYS392	6	14,867	4.036	1.48-8.78
YGATAH4	0	133	0.000	0-27.4	YGATAH4	19	7,709	2.464	1.49-3.84
DYS437	1	133	7.519	0.2-41.2	DYS437	12	10,101	1.188	0.61-2.07
DYS438	0	133	0.000	0-27.4	DYS438	3	10,122	2.964	0.06-0.87
DYS448	0	133	0.000	0-27.4	DYS448	11	6,678	1.647	0.83-2.95
DYS481	0	133	0.000	0-27.4	DYS481	3	314	9.554	1.98-2.77
DYS533	0	133	0.000	0-27.4	DYS533	0	314	0.000	0.00-11.68
DYS549	0	133	0.000	0-27.4	DYS549	1	314	3.185	0.087-17.62
DYS643	0	133	0.000	0-27.4	DYS643	1	314	3.185	0.081-17.62
DYS570	1	133	7.519	0.2-41.2	DYS570	7	314	22.900	9.00-45.39
DYS576	2	133	15.038	1.8-53.3	DYS576	6	317	19.110	7.04-41.12

Concordance Studies

Four hundred twenty eight samples previously typed with the PPY23 were also processed using the Y Filer kit as part of the concordance studies. Although thirty six samples (8.41%) were tagged as off ladder peaks or off marker range peaks, bin calling by Y filer yielded similar result with the PPY23. Allelic designation was made thru size comparison from the allelic ladder. All sizes concords with the result from PPY23. Table 6. summarizes the loci where OL and OMR were detected.

Table 6. summarizes the loci where allele calling becomes challenging and time consuming.

LOCUS	SAMPLE	Y FILER RESULT	ALLELE CALLING BY SIZE DETERMINATION	PPY23
DYS456	B4	OMR	9	9
	B5	OMR	9	9
	064-12-B	OL	19	19
	274-12-A681	OL	19	19
DYS458	PNPDB 94	OL	13	13
	PNPDB 150	OL	13	13
	PNPDB 155	OL	13	13
	PNPDB 163	OL	13	13
	230-11-H	OL	13	13
	227-12-A476	OL	13	13
	PNPDB38	OL	19	19
	PNPDB338	OL	15.1	15.1
DYS385 A/B	305-12-A876	15,OL	15,16.2	15,16.2
DYS635	PNPDB 78	OMR	28	28
	PNPDB 106	OMR	28	28
	PNPDB90	OL	27	27
	PNPDB91	OL	27	27
	PNPDB110	OL	27	27
	PNPDB319	OL	27	27
	076-12-A2	OL	27	27
Y GATA H4	PNPDB70	OL	14	14
	PNPDB81	OL	14	14
	PNPDB82	OL	14	14
	PNPDB126	OL	14	14
	PNPDB171	OL	14	14
	PNPDB290	OL	14	14
	PNPDB240	OL	14	14
	PNPDB340	OL	14	14
	230-11-I	OL	14	14

	230-D-G	OL	14	14
	230-D-H	OL	14	14
	128-12-F	OL	14	14
	351-12-A1155	OL	14	14
DYS448	082-13-A375	OL	14	14
	PNPDB 11	OL	19	19
DYS392	156-13-A695	OMR	15	15
	125-12-A	OL	13.1	13.1

Conclusion

The present study reports the generation of an expanded reference database consisting of 629 Y-STR profiles with representative samples from fourteen regions of the Philippines involving the analysis of 23 Y STR loci (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS481, DYS533, DYS549, DYS570, DYS576 and DYS643 DYS437, DYS448, DYS456, DYS458, DYS635, and Y-GATA-H4). Haplotype analysis of the Philippine population involving 23 Y STR markers system yields a haplotype diversity of 0.99992 with a standard error of 0.00230. This defines the presence of 612 haplotypes from the sample of 629 individuals. Out of the 612 haplotypes, 596 haplotypes were found to be unique, 15 haplotypes were found to be shared by two individuals and 1 haplotype was shared by three individuals. The discrimination capacity was calculated to be 97.30%. Eleven mutations observed from this study were all single step and is consistent with the strand slippage model for generating mutations for STR. The overall mutation rate calculated from this study is 3.6×10^{-3} (95% CI $(1.8-6.4) \times 10^{-3}$) and no bias for gains nor losses were recorded. Mutation data gathered from this study add up to the increasing data on Y STR mutation uploaded in the YHRD database. While diversity studies show the advantage of having more loci in discriminating Y STR profiles, concordance studies shows that both multiplex systems can be used to successfully type profiles and that the results of the Y Filer kit can be

compared directly with the results of PPY23. Information gathered in this study assists forensic interpretation by improving forensic probability calculation and understanding the diversity of Filipino male population.