Giuffrida, Stephanie M., <u>Establishing an STR allele frequency database for the</u> <u>Paraguayan population</u>. Master of Science (Biomedical Sciences), May, 2018, 35 pp., 7 tables, 1 figure, bibliography, 4 titles.

This project determined whether the loci included in the Athos PCR Directa para Identificacao Humana Amplification and AmpFlSTR® Identifiler® PCR Amplification Kits are suitable for establishment of an STR allele frequency database representative of the Paraguayan population. Allele frequencies for the 21 STR loci included in the Athos PCR Directa para Identificacao Humana Amplification Kit were calculated for 200 individuals in the Paraguayan population. Allele frequencies for the 15 STR loci included in the AmpFlSTR® Identifiler® PCR Amplification Kit were also calculated for 300 individuals in the Paraguayan population. Performance of these loci was determined by testing the population for Hardy-Weinberg Equilibrium, testing for Linkage Disequilibrium between the loci in the kits, and calculating Power of Discrimination, Power of Exclusion, and Mean Power of Exclusion. Comparability between the common loci included in the two PCR kits was also tested.

Results of these tests determined that the Athos PCR Directa para Identificacao Humana Amplification and AmpFlSTR® Identifiler® PCR Amplification Kits are both suitable for use in an STR allele frequency database. Additionally, the common loci contained in the two PCR kits are comparable, and the populations can be combined to create one database.

# ESTABLISHING AN STR ALLELE FREQUENCY DATABASE FOR THE PARAGUAYAN POPULATION

### THESIS

Presented to the Graduate Council of the Graduate School of Biomedical Sciences University of North Texas Health Science Center at Fort Worth In Partial Fulfillment of the Requirements

For the Degree of

### MASTER OF SCIENCE

By

Stephanie M. Giuffrida, B.S. Fort Worth, TX

May 2018

#### ACKNOWLEDGEMENTS

I would first like to thank my major professor, Dr. John Planz, for all of his advice, effort, and time provided during this project, and throughout my Graduate career at the UNT Health Science Center. A special thanks to my advisory committee, Dr. Joseph Warren, Dr. Nicole Phillips, and Dr. Rebecca Cunningham for dedicating their time and effort to meetings, reviews, and the overall success of my project. Next, I would like to thank the the Laboratorio Diaz Gill Medicina Laboratorial S.A. – Departamento de Identificación Genética (Paraguay) for allowing me the opportunity to complete this project with them, and for their dedicated time and contributions. Finally, I would like to thank my parents Tom and Tracy Giuffrida, and my family and friends for their unwavering love and support during my educational career. Thank you for always encouraging and supporting me to chase my dreams. I would not be as successful as I am today without you all.

### TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: RESEARCH DESIGN AND METHODOLOGY	5
2.1 Problem/ Hypothesis	5
2.2 Specimens Examined	5
2.3 DNA Extraction	7
2.4 Genetic Analysis	7
2.5 Statistical Analysis	9
CHAPTER 3: RESULTS	13
CHAPTER 4: DISCUSSION	18
4.1 Future Directions	19
APPENDIX	20
REFERENCES	27

## LIST OF TABLES

1.	Loci contained in the Athos PCR Directa papa Identificacao Humana
	Amplification and AmpFISTR <sup>®</sup> Identifiler <sup>®</sup> PCR Amplification
	kits9
2.	Allele frequencies and descriptive statistics of 21 STR loci included in
	the Athos® PCR amplification kit in the Paraguayan population13
3.	Allele frequencies and Exact Test for Hardy-Weinberg Equilibrium of
	21 STR loci included in the Athos® PCR amplification kit in the Paraguayan
	population calculated <i>poppr</i> and <i>pegas</i> in R14
4.	Allele frequencies and descriptive statistics of the 15 STR loci included in
	the Identifiler <sup>®</sup> PCR amplification kit in the Paraguayan population15
5.	Allele frequencies and Exact Test for Hardy-Weinberg Equilibrium of the
	15 STR loci included in the Identifiler PCR amplification kit in the Paraguayan
	population calculated <i>poppr</i> and <i>pegas</i> in R15
6.	Kolmogorov-Smirnov (KS) test results16
7.	Allele frequencies and descriptive statistics of the 15 STR loci included in
	the Athos® PCR and Identifiler PCR amplification kits17

# LIST OF FIGURES

1. Map of Paraguay depicting where analyzed samples were collected from......6

#### INTRODUCTION

Paraguay is a landlocked country located in Central South America, surrounded by Argentina, Bolivia, and Brazil. As of July 2017, Paraguay has a population of about seven million individuals. Much of the population resides in the Eastern and Southern-most parts of the country, with a small fraction of the population residing in the western plains of the Gran Chaco [1].

The main ethnic group of Paraguay is Mestizo, mixed Spanish and Amerindian, with ninety five percent of the population identifying as such. The remaining five percent identify as "other." Paraguay is a country of emigration, not immigration. Political instability, civil wars, and years of dictatorship have encouraged individuals to immigrate to neighboring countries. However, in the 20<sup>th</sup> century, minimal numbers of immigrants arrived in Paraguay from Lebanon, South Korea, Japan, and China. Additionally, since the 1960s, large numbers of Brazilian immigrants have arrived in Paraguay to work in agriculture [1].

Crime in Paraguay is generally non-violent; however, there is a prevalent use of knives and firearms, especially in street crimes. In a few instances, individuals have been injured or killed for defending themselves or refusing to cooperate. While men have reported to be victims of these crimes, women alone or in small groups tend to be the target. Additionally, armed robbery, burglary, car theft, and home invasion are common offenses in urban and rural areas [2]. Between October 2009 and February 2018, the Diaz Gill laboratory has analyzed 626 samples for the public prosecutor's office, including sexual assault, homicide, maternity, paternity, and identification samples. The largest amount of these samples, 270, pertains to sexual assault cases.

Additionally, the lab has analyzed about six thousand samples related to paternity/maternity relationship legal and non-legal cases (Diaz Gill).

If a biological sample is obtained from evidence at a crime, forensic analysts try to develop a genetic profile belonging to the contributor(s) of any DNA samples. If development of a genetic profile is successful, and there is a match between known individuals and the evidence, statistical evaluations are performed to determine the rarity of the genetic profile in a population and to determine the weight of the DNA evidence in any match or inclusion. Currently, Paraguay does not have an STR database representative of the country's population that can be used in human identification. The results of this study provide identity testing laboratories in Paraguay a validated population database for use in calculation of forensic and parentage evaluation statistics for criminal and civil investigations.

An STR database is developed by evaluating genetic data from a population sample for specific sets of genetic markers and statistical parameters. Data should be collected from random, non-related individuals for the most accurate representation of the desired population. Routinely, analysis of the genetic data includes calculation of the allele frequencies at each locus, testing the loci for Hardy-Weinberg Equilibrium (HWE) and testing for the potential presence of Linkage Disequilibrium between the loci in the STR typing panel. Linkage Disequilibrium is typically specific to the testing kit used.

First, the allele frequencies were calculated for each locus included in the testing kit. The allele frequency can be described as how often an allele appears at that specific locus. After calculating the allele frequencies at each locus, the software then calculated the Expected Heterozygosity and Observed Heterozygosity. Expected Heterozygosity (H<sub>e</sub>) is defined as the expected genotype frequencies of the heterozygotes in the population. It is calculated by the

2

software, based on the genotype data entered. In contrast, the Observed Heterozygosity  $(H_0)$  is the actual genotype frequencies of the heterozygotes observed in the population.

Next, the software tested for Hardy-Weinberg Equilibrium in the population. Hardy-Weinberg Equilibrium is a mathematical model which allows an individual to predict expected genotype frequencies based on the allele frequencies determined for the loci tested, if a population is following the general patterns of Mendelian inheritance and meets other model assumptions. These assumptions include that: the organism in question is diploid, mating is random, reproduction is sexual, generations are non-overlapping, population size is large, migration is negligible, and mutation can be ignored. If these assumptions can be met, allele frequencies in a population will not change substantially from generation to generation [3]. These allele frequencies can be used to estimate the probability of observing any genotype generated from the STR loci tested in human identification.

In genetics, Linkage Disequilibrium is the non-random association of alleles between different loci in a population. In contrast, if sets of loci in a population exhibit Linkage Equilibrium, individual genomes reflect a random association of alleles in offspring. This is important, because if the genetic loci demonstrate Linkage Equilibrium, the genetic composition of the population can be fully represented by the allele frequencies of the loci tested. If loci are not in Linkage Equilibrium, they cannot be included in forensic statistic calculations using the product rule as they do not reflect independent variables [4].

Statistical parameters relevant to identity testing that are routinely calculated for forensic population databases include Power of Discrimination (PD), Power of Exclusion, and Mean Power of Exclusion. Power of Discrimination is the probability that two individuals in a population have identical genotypes. In contrast, the Power of Exclusion, first described by

3

Fisher in 1951, estimates the probability that two randomly selected individuals in a population do not have identical genotypes [5]. Mean Power of Exclusion, as described by Brenner and Morris, is most commonly calculated for parentage testing purposes. Mean Power of Exclusion provides the average power of exclusion for each locus between an offspring and potential parental alleles [6].

In this project, two data analysis approaches were utilized to calculate the statistical parameters mentioned above. The first is Genetic Data Analysis, GDA, created by Lewis and Zaykin [7]. The second approach utilized statistical packages developed in R, an environment and language for data analysis, created by the R Foundation for Statistical Computing to which a multitude of contributors provide software tools for various analyses and functions [8]. Both approaches were utilized, and the results were compared to explore more current approaches for allele frequency database analysis.

#### RESEARCH DESIGN AND METHODOLOGY

#### *i. Problem/Hypothesis*

There are three specific aims of this project:

- 1. Establish an allele frequency database for the country of Paraguay.
- 2. Determine if the allele frequencies calculated at the shared loci are comparable between the two PCR amplification kits: Athos (Brazil) and Identifiler®.
- 3. Create a pipeline analysis method for the creation of an STR databases.

There are multiple hypotheses for this project, as there are multiple tests conducted on the data. The first null hypothesis ( $H_0$ ) is that the population is in Hardy-Weinberg Equilibrium. Next, is that the population is in Linkage Equilibrium, and finally, the third null hypothesis is that the common loci included in the two PCR kits are concordant with one another, as determined by the KS test.

In contrast, the alternative hypotheses (H<sub>a</sub>) are as follows: the population is not in Hardy-Weinberg Equilibrium, the population does not exhibit Linkage Equilibrium, and the common loci included in the two PCR kits are not concordant with one another.

#### *ii. Specimens Examined*

A total of five hundred unrelated individuals of the Paraguayan population were sampled during routine parentage and criminal casework protocols. Two hundred samples were analyzed using the Athos PCR Directa para Identificacao Humana Kit (Brazil) were collected from seventeen departments within the country; Asunción, Alto Paraná, Amambay, Boquerón, Caaguazú, Caazapá, Canindeyú, Central, Concepción, Cordillera, Guairá, Itapúa, Misiones, Ñeembucú, Paraguarí, Presidente Hayes, San Pedro. The remaining three hundred samples, to be analyzed using the AmpFlSTR® Identifiler® PCR Amplification kit (Applied Biosystems®), were collected from sixteen of those seventeen departments, with the exception of Amambay (Figure 1).



Figure 1. Map of Paraguay depicting where the analyzed samples were collected from. Circles represent the locations where the Athos kit samples were collected from. Squares represent the locations where the Identifiler kit samples were collected from (adapted from [9]).

Buccal swabs, peripheral blood in EDTA and peripheral blood stored on FTA or filter paper cards were collected by Dr. Karen Nielsen of Diaz Gill: Medicina Labrotorial S.A., located in Asunción, Paraguay. De-identified genetic profiles were provided by Dr. Nielsen for use in this project. Working with these data was approved under Institutional Review Board (IRB) protocol 2010-106.

#### *iii.* DNA Extraction

DNA extraction was performed at the Diaz Gill laboratory in Paraguay. Liquid samples and buccal swab samples were extracted using commercial PrepFiler<sup>™</sup> BTA Forensic DNA Extraction Kit (ThermoFisher Scientific, USA), Ultraclean® BloodSpin® DNA Isolation Kit (MO BIO Laboratories, Inc. Carlsbad, CA), DNA IQ Casework Pro Kit for Maxwell<sup>®</sup> 16 Kit (Promega Corp., Madison, WI) and Chelex<sup>®</sup> Resin (Bio-Rad Laboratories, Hercules, CA). Direct PCR was performed on FTA substrates according to the manufacturer's instructions.

#### *iv. Genetic Analysis*

Databases were developed for two commercially available PCR amplification kits which were used to type a total of 500 individuals in the sample sets. For 300 samples, 15 STR genetic markers were amplified with the commercial AmpFlSTR® Identifiler® PCR Amplification Kit (Applied Biosystems®), following the manufacturer's recommended PCR protocol in the Veriti Thermal Cycler (Applied Biosystems®, Foster City, CA). The detection of the amplified product was performed in the genetic sequencer ABI 3130x1 (Applied Biosystems®, Foster City, CA) and the reading and allele assignment with GeneMapper® software v1.2.1 (ThermoFisher Scientific, USA). The AmpFlSTR® Identifiler® PCR Amplification kit (Applied Biosystems®) contains 15 autosomal loci, plus amelogenin. The 15 autosomal loci include CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX, and vWA, which are all included in the Athos kit (Brazil) as well (Table 1) [10].

For 200 samples, 21 STR genetic markers were amplified with the commercial Athos PCR Directa para Identificacao Humana Amplification Kit. The detection of the amplified product was performed in the ABI 3130xl Genetic Analyzer (Applied Biosystems<sup>®</sup>, Foster City, CA, USA) and the allele assignment with GeneMapper® software v1.2.1 (ThermoFisher Scientific, USA). The Athos PCR kit contains 21 autosomal loci, plus amelogenin. These 21 autosomal loci include D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1237, D1S1656, D2S1338, Penta E, Penta D, and HUMF13B [11]. This kit is an improvement from the Identifiler<sup>®</sup> kit, as an increased number of loci provides an increased discriminating power during genetic analysis. **Table 1.** Loci contained in the Athos<sup>®</sup> PCR Directa papa Identificacao Humana Amplification and AmpFISTR<sup>®</sup> Identifiler<sup>®</sup> PCR Amplification kits. "X" denotes loci that are included in the kit.

Loci	Athos <sup>®</sup> PCR Directa para Identificacao Humana	AmpFlSTR <sup>®</sup> Identifiler <sup>®</sup> PCR Amplification kit
	Amplification Kit	
D3S1358	X	Х
vWA	X	Х
D168539	Х	Х
CSF1PO	Х	Х
TPOX	X	Х
D8S1179	X	Х
D21S11	Х	Х
D18S51	X	Х
D19S433	X	Х
TH01	X	Х
FGA	X	Х
D22S1045	X	
D5S818	X	Х
D13S317	X	Х
D7S820	Х	Х
D10S1237	X	
D1S1656	X	
D2S1338	Х	Х
Penta E	X	
Penta D	Х	
HUMF13B	X	

#### v. Statistical Analysis

The Athos PCR Amplification Kit data and the AmpFlSTR<sup>®</sup> Identifiler<sup>®</sup> PCR Amplification Kit (Applied Biosystems<sup>®</sup>) data were formatted in Microsoft Excel and Notepad for import into GDA software [7]. GDA was used to calculate the observed allele frequencies at each locus, as well as testing for Hardy-Weinberg Equilibrium and potential linkage disequilibrium between the loci in the kits. After calculating the allele frequencies, the GDA software calculated the estimate of expected heterozygote/homozygote frequencies, based on the observed allele frequencies, and compared these to the observed genotype distributions. Then, Fisher's Exact Test was utilized to determine whether the population exhibited Hardy-Weinberg Equilibrium [7]. Fisher's Exact Test predicts the probabilities of possible genotypes based on the alleles present in the population. When sample size is small, and/or some allele frequencies are low, the Exact Test is preferred over a Chi-square ( $\chi^2$ ) approaches, or likelihood ratio statistic  $G^2$ , as they can often result in false acceptance or rejection of Hardy-Weinberg Equilibrium when population size is small [12].

The PowerStats v1.2 software (Promega) was utilized to calculate the Power of Discrimination and Mean Power of Exclusion for each locus in the population [5]. Power of Discrimination is calculated as one minus the sum of the squares of the observed phenotype frequencies in the observed set of data. The Fisher's Power of Exclusion was also calculated at each locus in the population using an Excel spreadsheet developed in-house. Power of Exclusion is calculated as one minus the sum of the squares of the expected phenotype frequencies in the observed set of data [13]. The Power of Exclusion provides the percentage of the population that could be excluded as the source of the observed genetic DNA.

To determine if the allele frequency distributions for the overlapping loci between the two PCR amplification kits are comparable, the Kolmogorov-Smirnov (KS) test was used [14]. The KS test compared the distribution of allele frequencies between each locus in the kit, and provided a *p*-value, as well as a Kolmogorov Distribution value, *D*. The *p*-value determines if the allele frequency distributions differ significantly at the loci between the population data sets. If the *p*-value is less than the significant value for the test, the null hypothesis, that the common

loci between the two PCR kits are comparable with one another, should be rejected, indicating that there is a difference between the datasets. The D statistic calculated is a measure of the deviation between the two datasets. If you were to visualize the two sets of data on a curve, the *D* statistic would be the greatest vertical distance between the two curves at any point [15].

Although GDA can provide descriptive statistics and tests for disequilibrium, as well as population comparisons and genetic structure, it is an old program and designed to run on earlier versions of Microsoft Windows<sup>™</sup>. Additionally, it requires a specific input file format. Software packages in the R language were also utilized to evaluate the aspects of genetic data. The values obtained from the R packages were combined with, and compared to, the information obtained from GDA to determine the best methods to use and in summarizing the overall project results. The goal was to develop a pipeline script that can conduct all the analyses in a single run.

In the R language analysis, a variety of packages, such as *poppr* and *pegas*, were utilized to analyze the genetic data. The R package *poppr* is designed for the analysis of sexual, admixed, and/or clonal populations. *poppr* has specific functions that were designed to easily import data from excel and read data from diploid markers [16]. The package was used to count alleles per locus in the population, and in turn provided the same allele frequencies that were also calculated in GDA.

The R package *pegas* was utilized to determine if the population follows Hardy-Weinberg Equilibrium criteria. The package can determine HWE by two methods; the classical chi-squared ( $\chi^2$ ) method, based on the expected genotype frequencies, or the exact test based on Monte Carlo permutations of alleles [17]. This project utilized the Fisher's exact test because of the sample size available for this study.

11

Additional packages were explored to test for Linkage Disequilibrium in the populations, however, none of the available packages are designed for use on multiallelic markers. Some of these packages include *LDtest*, *genetics*, and *pegas*.

#### RESULTS

Five hundred samples in the Paraguayan population were evaluated for parameters typical of an STR database for use in human identification; Expected Heterozygosity and Observed Heterozygosity, Hardy-Weinberg Equilibrium, Linkage Disequilibrium, Power of Discrimination, Power of Exclusion, and Mean Power of Exclusion. Two PCR amplification kits, Athos PCR Directa para Identificacao Humana (Brazil) and AmpFISTR<sup>®</sup> Identifiler<sup>®</sup> PCR Amplification Kit (Applied Biosystems<sup>®</sup>), were used to generate population data for STR loci.

Allele frequencies were calculated for 200 individuals at 21 STR loci in the Athos<sup>®</sup> PCR Directa para Identificacao Humana amplification kit (Brazil) (Table 2). Expected Heterozygosity (H<sub>e</sub>) and Observed Heterozygosity (H<sub>o</sub>) were also calculated. Fisher's Exact Test was used to determine whether the loci exhibit Hardy-Weinberg Equilibrium. Exact Test probabilities are based on 3200 shufflings. Shuffling randomizes the loci in the sample set to determine the probability of observing that data set given that there is no disequilibrium between the loci. One significant deviation from Hardy-Weinberg Equilibrium was detected. A locus can be said to deviate from Hardy-Weinberg Equilibrium when the *P* value calculated is less than the significance value at which the HWE hypothesis would be rejected. A significance value of  $\alpha$ = 0.05 was used in this study. The *p*-value calculated for the D13S317 locus is 0.028125, which is less than the significance value of 0.05. This is not unusual, as it is common for the likelihood of error to occur as the number of independent tests increases [18]

To test for significant departure within the entire suite of loci, one must look closer at the individual significance of the locus opposed to the overall significance. A Bonferroni correction

was applied to the *p*-value of the individual test, in order to maintain the overall *P* value of  $\alpha$ = 0.05 [18]. With the Bonferroni correction, the adjusted *p*-value becomes  $\alpha$ = 0.002, reducing the threshold probability for considering a departure from HWE. Employing this corrected significant value, D13S317 does not deviate from Hardy-Weinberg Equilibrium. In its entirety, the loci included in the Athos<sup>®</sup> PCR Directa para Identificacao Humana amplification kit (Brazil) exhibit Hardy-Weinberg Equilibrium, and can be used for inferring the genotype probabilities based on the allele frequency data that are reported here.

The R language was also used to determine whether the population deviated from Hardy-Weinberg Equilibrium (Table 3). As expected, the calculated *p*-values differed slightly from those calculated in GDA, and this occurs because the 3200 shufflings in the test are not exactly the same each time the test is conducted. For the loci tested in the R platform, all calculated *p*-values were above the significant value of  $\alpha$ = 0.05, indicating that the population does meet Hardy-Weinberg Equilibrium conditions and can be used in human identity and parentage testing calculations.

The 21 autosomal loci included in the Athos<sup>®</sup> PCR Directa para Identificacao Humana amplification kit (Brazil) were also tested for Linkage Disequilibrium in GDA, determined by the calculated allele frequencies. Results are based on 210 pairwise comparisons. To maintain the overall significant value of  $\alpha$ = 0.05 for the panel of loci, each individual locus should have had a *p*-value of at least  $\alpha$ = 0.0002442. There were no loci with *p*-values below this number, so the entire panel was determined to exhibit Linkage Equilibrium and can therefore be included in the database for human identification and parentage testing.

Statistics commonly calculated in human identification and parentage testing, such as Power of Discrimination, Power of Exclusion, and Mean Power of Exclusion, were also calculated (Table 2). Power of Discrimination and Power of Exclusion indicate the probability that two individuals will have the same or different genotypes. The Power of Discrimination (PD) indicates how individualizing the loci are. The Combined Power of Discrimination for this suite of loci was greater than 99.999999%. The values calculated in this project are comparable to the calculated Power of Discrimination values for published population databases in the countries surrounding Paraguay, such as Brazil [19] and Argentina [20]. In contrast to the Power of Discrimination, the Power of Exclusion (PE) indicates how powerful the loci are at excluding particular genotypes. The Combined Power of Exclusion for this set of loci is also greater than 99.9999999 %. In regard to both of these statistics, the greater the value, the more powerful the loci are.

For the remaining 300 samples, which were analyzed using the AmpFlSTR<sup>®</sup> Identifiler<sup>®</sup> PCR Amplification Kit (Applied Biosystems<sup>®</sup>), allele frequencies were calculated for 15 autosomal loci included in the kit (Table 4). As with the 200 samples analyzed using the Athos PCR Directa para Identificacao Humana amplification kit (Brazil), Expected Heterozygosity (H<sub>e</sub>) and Observed Heterozygosity (H<sub>o</sub>) were both calculated. Fisher's Exact Test was used to determine whether the loci exhibited Hardy-Weinberg Equilibrium. Again, results were based on 3200 shufflings. All *p*-values calculated were greater than  $\alpha$ = 0.05, and no significant deviations from Hardy-Weinberg Equilibrium were observed.

As previously mentioned, the R language was also used to calculate the allele frequencies and Hardy-Weinberg Equilibrium *P*-values for the fifteen loci (Table 5). Again, the *p*-values calculated differed slightly from the values calculated in the GDA software, but this is expected. For the loci tested in the R language, all calculated *p*-values were above the significant value of  $\alpha$ = 0.05, indicating that the population does meet Hardy-Weinberg Equilibrium conditions and can be used in human identity and parentage testing calculations. This suite of loci was also tested from Linkage Disequilibrium to determine if the loci can be used in statistical calculations for human identification and parentage testing. After 105 pairwise comparisons, there were no loci that exhibited Linkage Disequilibrium. All of the calculated *p*-values were above the significant value of  $\alpha$ = 0.0004, and therefore maintaining the overall significant value of  $\alpha$ = 0.05.

Power of Discrimination, Power of Exclusion, and Mean Power of Exclusion, were also calculated (Table 4). The Combined Power of Discrimination for this suite of loci was greater than 99.999999%, and the Combined Power of Exclusion was greater than 99.999999%. The values calculated are comparable to the calculated Power of Discrimination values for published population databases in the countries surrounding Paraguay, such as Brazil [19] and Argentina [20].

While both suites of STR loci exhibit Hardy-Weinberg Equilibrium and Linkage Equilibrium separately, the two PCR kits share fifteen loci, so they were tested for deviations in their allele frequencies using the Kolmogorov-Smirnov (KS) test (Table 6). The KS test produces two values, the *P*-value, and the *D* statistic [14]. The allele frequencies for each locus were compared to one another, and the *p*-value indicated if the numbers differed significantly from one kit to the other. Had the value been very small, this would have indicated that the locus was not concordant between the two kits. In this study, all of the *p*-values were about 0.600 or greater, indicating that the allele frequencies per locus do not differ significantly between the two kits. Additionally, the *D* statistic, which is the deviation between the allele frequencies, is about 0.300 or lower per locus. This indicates, again, that the allele frequencies do not differ significantly per locus between the two kits, and they can be combined for use in one database [15].

The two populations, two hundred samples tested with the Athos<sup>®</sup> PCR kit and three hundred samples with Identifiler<sup>®</sup> PCR kit, were combined for fifteen shared loci. The allele frequencies were then calculated for the single population using the fifteen loci common to both kits (Table 7). As with the other data sets, the population was also tested for Hardy-Weinberg Equilibrium, Linkage Disequilibrium, and the common parentage testing statistics.

When the two datasets were combined, there were two loci that did not exhibit Hardy-Weinberg Equilibrium. Results are based on 3200 shufflings. Locus D5S818 had a calculated *P*-value of  $\alpha$ = 0.008, which was below the significant value of  $\alpha$ = 0.05. Locus CSF1PO, which also exhibited departure from Hardy-Weinberg Equilibrium, had a calculated *p*-value of  $\alpha$ = 0.048. As with loci included in the Athos PCR kit, a Bonferroni correction was applied to the individual loci that did not pass tests for Hardy-Weinberg equilibrium. The adjusted *p*-value, determined by the Bonferroni correction, was  $\alpha$ = 0.003. After the correction, both individual loci were greater than the significant value, and were determined to exhibit Hardy-Weinberg Equilibrium, which indicates that they can be used in statistical calculations for human identity and parentage testing.

After testing the suite of loci for Linkage Disequilibrium, it was determined that none of the loci exhibited Linkage Disequilibrium. While some calculated *p*-values fell below the significant value of  $\alpha$ = 0.05, none of the individual locus *p*-values were lower than  $\alpha$ = 0.0004, which maintains the overall  $\alpha$ = 0.05. With this adjusted *p*-value, all of the individual locus significant values passed the Linage Disequilibrium test. This indicates that the loci can be used for statistical calculations for human identity and parentage testing.

#### DISCUSSION

This study evaluated allele frequency data for 21 STR loci and five hundred individuals of the Paraguayan population to determine whether the population could be used to create the first allele database for the country to use in human identity and parentage testing. Currently, the laboratories in Paraguay use a published database originating from Columbia which is not representative of the country's population.

After calculating the allele frequencies per locus, the Expected Heterozygosity (H<sub>e</sub>) and Observed Heterozygosity (H<sub>o</sub>) were calculated and used with the Fisher's Exact Test to determine whether the population is in Hardy-Weinberg Equilibrium. All three tests for departure from Hardy-Weinberg Equilibrium and Linkage Disequilibrium demonstrated that the population does in fact exhibit Hardy-Weinberg Equilibrium and Linkage Equilibrium and can be used for statistical calculation if human identity and parentage testing. In fact, the two populations analyzed using the Athos<sup>®</sup> PCR Directa para Identificacao Humana Amplification kit (Brazil) and AmpFISTR<sup>®</sup> Identifiler<sup>®</sup> PCR Amplification Kit (Applied Biosystems<sup>®</sup>) can be used as two separate databases or can be combined to create one database containing their fifteen shared loci (Table 1).

If the Diaz Gill: Medicina Labrotorial S.A. (Paraguay) plans on using both the Athos<sup>®</sup> (Brazil) and Identifiler<sup>®</sup> (Applied Biosystems<sup>®</sup>) PCR amplification kits for their casework, it would be beneficial to have both population databases available for use in their human identification and parentage testing statistical calculations. This is because utilizing only one database that contains the shared fifteen loci eliminates six descriptive loci that are included in

only the Athos<sup>®</sup> PCR Directa para Identificacao Humana Amplification kit (Brazil). The Diaz Gill: Medicina Labrotorial S.A. (Paraguay) lab should also consider keeping the combined database as well because while it does only contain fifteen loci, it will contain five hundred samples in total, increasing the power of the database.

#### Future Directions:

Establishment of an STR allele frequency database depends on multiple analysis methods, such as GDA, the R language, and several excel spreadsheets. In order to finalize a complete pipeline analysis for the creation of an allele frequency database, it would be beneficial to create a script which can easily calculate allele frequencies, test for Hardy-Weinberg Equilibrium, Linkage Disequilibrium between loci in an STR amplification kit, and calculate Power of Discrimination, Power of Exclusion, and Mean Power of Exclusion. While the calculation of allele frequencies and the tests for Hardy-Weinberg equilibrium were simple and efficient, there are no packages available to test for Linkage Disequilibrium among multiallelic markers. Moving forward, a script should be developed that can test for Linkage Disequilibrium, compare allele frequencies between loci in a PCR kit, and calculate other statistics. APPENDIX

Table 2: Allele frequencies and descriptive statistics of 21 STR loci included in the Athos® PCR amplification kit in the Paraguayan population

Allele	HUMTPOX	HUMvWA	D13S317	D18S51	Penta E	D5S818	D21S11	HUMCSF1PO	D10S1237	D2S1338	D3S1358	D16S539	D7\$820	D1S1656	HUMFIBRA_FGA	D8S1179	D19S433	HUMTH01	HUMF13B	D22S1045	Penta D
	N:200	N:200	N:200	N:200	N:200	N:200	N:200	N:200	N:200	N:200	N: 200	N: 200	N: 200	N: 200	N:200	N:200	N:200	N:200	N:200	N:200	N:200
2.2																					0.006
5				0.003	0.020																0.003
6				-	-	0.075												0.063	0.063		-
7				-	0.115	0.075		0.003					0.005					0.253	0.013		0.003
8	0.428		0.070	-	0.033	0.013		-				0.013	0.110			0.023		0.090	0.165		0.015
9	0.078		0.150	-	0.015	0.030		0.018				0.150	0.135			0.003		0.133	0.338		0.220
9.5	0.059		0.050	0.009	0.049	0.059		- 0.233	0.003			0 112	- 0.230	0.003		-	0.008	0.216	- 0.423		0.233
10	0.038	0.003	0.000	0.008	0.113	0.038		0.233	0.003			0.113	0.230	0.005		0.095	0.003	0.010	0.423	0.043	0.235
12	0.108	-	0.280	0.123	0.138	0.215		0.425	-			0.280	0.223	0.085		0.158	-			-	0.180
12.2		-	-	-	-	-		-				-	-			-	0.020			-	-
13		-	0.178	0.090	0.105	0.148		0.065			0.013	0.015	0.020	0.080		0.283	0.243			0.006	0.123
13.2		-	-	-	-	-		-	-		-	-	-	-		-	0.048			-	-
14		0.055	0.070	0.240	0.090	0.013		0.008	-		0.050	0.015	0.005	0.140		0.230	0.278			0.028	0.088
14.2		-	-	-	-	-			-		-	-	-	-		-	0.040			-	-
15		0.093	0.003	0.123	0.120	0.005			0.003		-	0.015	0.003	0.110		0.093	0.160			0.493	0.003
15.2		-		-	-				-		-			-			0.063			-	-
15.3		-		-	-				-		-			0.020		-	-			-	-
16		0.290		0.143	0.070				0.003	0.038	0.253			0.178		0.018	0.025			0.365	-
16.2		-		-	-				-	-	-			-		-	0.015			-	-
10.5		0.915		0 1 1 2	0.055				- 0.003	- 0.220	0 193			0.040	0.003	-	0.003			-	0.005
17.3		0.313		0.115	0.055				0.003	0.220	0.165			0.040	0.005	0.003	0.005			0.005	0.005
18		0 185		0.073	0.020				0.003	0.115	0 103			0.133	0.020					-	
18.2		-		-	-				-	-	-			-	0.003					-	
18.3		-			-				-					0.065	-					-	
19		0.058		0.033	0.020				0.205	0.123	0.015			-	0.050					-	
19.3		-		-	-				0.003					0.013	-					-	
20		-		0.038	0.018				0.250	0.123					0.080					-	
21		0.003		0.003	0.020				0.115	0.033					0.155					-	
21.3				-	-				0.003						-					-	
22				0.005	0.003				0.095	0.073					0.128					0.003	
22.2									-	0 125					0.008					-	
23									0.005	0.125					0.103						
23.2				0.003					0.052	0 103					0.163					-	
25									0.003	0.038					0.165					-	
25.2							0.005								-					-	
26							-			0.013					0.088					-	
27							0.013								0.013					-	
28							0.143								0.010					0.003	
29							0.218								0.003						
30							0.290								-						
30.2							0.028								-						
31							0.045								-						
51.2							0.065								-						
32 3							0.010														
34.2							0.005								-						
45															0.003						
Here	0.689	0.769	0.817	0.866	0.911	0.726	0.824	0.700	0.852	0.875	0.743	0.792	0.796	0.880	0.880	0.862	0.823	0.777	0.678	0.619	0.829
Hobs	0.695	0.720	0.810	0.830	0.860	0.740	0.835	0.690	0.860	0.900	0.710	0.790	0.810	0.890	0.880	0.810	0.810	0.815	0.685	0.685	0.820
P	0.831	0.243	0.028	0.545	0.072	0.073	0.818	0.740	0.569	0.452	0.810	0.377	0.755	0.628	0.512	0.978	0.591	0.268	0.282	0.563	0.707
PE	0.850	0.910	0.941	0.967	0.984	0.889	0.948	0.859	0.972	0.971	0.981	0.951	0.926	0.973	0.973	0.941	0.952	0.967	0.836	0.784	0.947
PD	0.848	0.910	0.929	0.965	0.980	0.883	0.944	0.858	0.958	0.967	0.891	0.919	0.923	0.969	0.969	0.941	0.944	0.908	0.831	0.756	0.944
MPE	0.421	0.460	0.618	0.656	0.715	0.493	0.666	0.413	0.715	0.795	0.444	0.581	0.618	0.775	0.775	0.618	0.618	0.627	0.405	0.405	0.637

Hege- expected heterozygosity, Hear- observed heterozygosity, P- Hardy-Weinberg exact test based on 3200 shufflings, N- number of individuals typed, PE- Power of Exclusion, PD- Power of Discrimination, MPE- Mean power of Exclusion

## Table 3: Allele frequencies and Exact Test for Hardy-Weinberg Equilibrium of 21 STR loci included in the Athos® PCR amplification

$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	Allele	HUMTPOX N: 200	HUMvWA N: 200	D13S317 N: 200	D18S51 N: 200	Penta E N: 200	D5S818 N: 200	D21S11 N: 200	HUMCSF1P N: 200	C D10S1237 N: 200	D2S1338 N: 200	D3S1358 N: 200	D16S539 N: 200	D7S820 N: 200	D1S1656 N: 200	MFIBRA_F( N: 200	D8S1179 N: 200	D19S433 N: 200	HUMTH01 N: 200	HUMF13B N: 200	D22S1045 N: 200	Penta D N: 200
5       0.013       0.013       0.013       0.003       0.023       0.0	2.2																					0.006
1	5				0.003	0.020																0.003
1       0       0.03       0	6				-	-													0.063	0.063		-
b         0.023         0.013         0.0	7	0.420		0.070	-	0.115	0.075		0.003				0.012	0.005			0.022		0.253	0.013		0.003
9         00/0         01/0         0.00         0100         0000         0100         0100         0100         0000         01000         0100         0100         0	8	0.428		0.070	-	0.033	0.013		-				0.013	0.110			0.023		0.090	0.100		0.015
no         0.050         0.050         0.068         0.048         0.233         0.033          0.113         0.230         0.033          0.038         0.038         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.058         0.051<	93	0.078		0.150	-	0.015	0.030		0.018				0.150	0.155			0.005		0.133	0.338		0.220
11       0.339       0.033       0.200       0.081       0.113       0.443       0.248       0.037       0.233       0.237       0.203       0.035       0.016       0.015       0.015       0.015       0.015       0.015       0.016       0.015       0.	10	0.058		0.050	0.008	0.048	0.058		0.233	0.003			0.113	0.230	0.003		0.098	0.008	0.010	0.423		0.233
112       0.108       -       0.209       0.139       0.139       0.139       0.139       -       -       0.100       0.209       0.232       0.055       0.013       0.020       0.213       0.213       0.213       0.213       0.015       0.020       0.201       0.233       0.213       0.021	11	0.330	0.003	0.200	0.008	0.113	0.445		0.248	0.003			0.253	0.270	0.025		0.095	0.003			0.043	0.125
112	12	0.108	-	0.280	0.123	0.138	0.215		0.425	-			0.280	0.223	0.085		0.158	-			-	0.180
13.2        0.178       0.080       0.165       0.148       0.055       0.07       0.248       0.248       0.248       0.026       0.125         144       0.055       0.070       0.240       0.090       0.013       0.005       0.015       0.005       0.014       0.230       0.278       0.028       0.028       0.028       0.028       0.029       0.0278       0.028       0.029       0.0278       0.028       0.028       0.028       0.029       0.040       0.028       0.028       0.029       0.040       0.049       0.040       0.049       0.008       0.029       1.6       0.278       0.049       0.049       0.005       0.015<	12.2		-	-	-	-	-		-	-			-	-	-		-	0.020			-	-
13.2       -	13		-	0.178	0.090	0.105	0.148		0.065	-		0.013	0.015	0.020	0.080		0.283	0.243			0.006	0.123
142       0.035       0.070       0.240       0.090       0.013       0.005       0.010       0.005       0.010       0.021       0.022       0.028       0.028       0.021         15       0.093       0.03       0.123       0.120       0.005       0.005       0.015       0.015       0.003       0.101       0.099       0.160       0.449       0.043         153           0.015       0.003       0.016       0.093       0.0163        0.015       0.0163       0.015       0.0163       0.015       0.013       0.015       0.0163          0.015       0.013       0.015       0.013       0.015       0.013       0.015       0.013       0.015       0.013       0.015       0.013       0.015       0.013	13.2		-	-	-	-	-		-	-		-	-	-	-		-	0.048			-	-
112             0.015       0.033       0.109       0.049       0.049       0.049         115       0.093       0.123       0.003       0.100       0.093       0.100       0.093       0.100       0.049       0	14		0.055	0.070	0.240	0.090	0.013		0.008	-		0.050	0.015	0.005	0.140		0.230	0.278			0.028	0.088
152       -       -       -       -       0.035       0.049       0.013       0.055       0.066       0.057       0.066       0.057       0.065       0.013	14.2		-	-	-	-	-			-		-	-	-	-		-	0.040			-	-
11.3       -       -       -       -       -       -       0.003       0.0143       0.070       0.003       0.0253       0.0175       0.018       0.015       0.035       0.035       0.0175       0.015 <td< td=""><td>15.2</td><td></td><td>0.093</td><td>0.003</td><td>0.123</td><td>0.120</td><td>0.005</td><td></td><td></td><td>0.003</td><td></td><td>-</td><td>0.015</td><td>0.003</td><td>0.110</td><td></td><td>0.093</td><td>0.160</td><td></td><td></td><td>0.493</td><td>0.003</td></td<>	15.2		0.093	0.003	0.123	0.120	0.005			0.003		-	0.015	0.003	0.110		0.093	0.160			0.493	0.003
16         0.280         0.143         0.070         0.063         0.025         0.170         0.018         0.025         0.365         1           16         -	15.2		-		-	-				-		-			0.020		-	0.003			-	-
16.2 <td>16</td> <td></td> <td>0 290</td> <td></td> <td>0 143</td> <td>0 0 7 0</td> <td></td> <td></td> <td></td> <td>0.003</td> <td>0.038</td> <td>0 253</td> <td></td> <td></td> <td>0.178</td> <td></td> <td>0.018</td> <td>0.025</td> <td></td> <td></td> <td>0 365</td> <td></td>	16		0 290		0 143	0 0 7 0				0.003	0.038	0 253			0.178		0.018	0.025			0 365	
163 <td>16.2</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td>-</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>0.015</td> <td></td> <td></td> <td>-</td> <td>-</td>	16.2										-	-						0.015			-	-
17       0.315       0.113       0.055       0.033       0.049       0.003       0.003       0.003       0.003       0.003       0.003       0.005       0.	16.3		-								-	-			0.040		-	0.003			-	
17.3       -       -       -       -       -       0.033       0.115       0.030       0.115       0.030       0.115       0.030       0.115       0.031       0.115       0.031       0.115       0.031       -       -       -       -       -       -       -       -       0.010       0.020       0.013       0.010       0.020       -       -       -       -       -       -       0.003       0.115       0.031       -       -       -       -       -       -       -       0.010       0.020       -       -       -       -       -       -       0.031       -       -       -       -       -       -       -       0.033       -       -       -       -       -       -       -       0.033       - <td>17</td> <td></td> <td>0.315</td> <td></td> <td>0.113</td> <td>0.055</td> <td></td> <td></td> <td></td> <td>0.003</td> <td>0.220</td> <td>0.183</td> <td></td> <td></td> <td>0.040</td> <td>0.003</td> <td>0.003</td> <td>0.003</td> <td></td> <td></td> <td>0.063</td> <td>0.005</td>	17		0.315		0.113	0.055				0.003	0.220	0.183			0.040	0.003	0.003	0.003			0.063	0.005
18       0.185       0.073       0.020       0.003       0.115       0.103       0.010       0.020       -         18.3       -       -       -       -       -       -       0.005       -       0.005       -       -       -       -       -       -       -       0.005       0.010       0.020       -       -       -       -       -       -       0.005       0.013       0.010       0.020       -       -       -       -       -       0.005       0.010       0.010       0.020       -       -       0.005       -	17.3		-		-	-				-	-	-			0.193	-					-	
18.2       -       -       -       -       -       -       0.003       -       -       -         19       0.058       0.033       0.020       0.025       0.123       0.015       -       0.050       - <td>18</td> <td></td> <td>0.185</td> <td></td> <td>0.073</td> <td>0.020</td> <td></td> <td></td> <td></td> <td>0.003</td> <td>0.115</td> <td>0.103</td> <td></td> <td></td> <td>0.010</td> <td>0.020</td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td></td>	18		0.185		0.073	0.020				0.003	0.115	0.103			0.010	0.020					-	
18.3 <td< td=""><td>18.2</td><td></td><td>-</td><td></td><td>-</td><td>-</td><td></td><td></td><td></td><td>-</td><td>-</td><td>-</td><td></td><td></td><td>-</td><td>0.003</td><td></td><td></td><td></td><td></td><td>-</td><td></td></td<>	18.2		-		-	-				-	-	-			-	0.003					-	
19       0.038       0.038       0.039       0.039       0.039       0.039       0.030       0.030       0.030       0.030       0.030       0.031       0.033       0.	18.3		-		-	-				-	-	-			0.065	-					-	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	10.2		0.058		0.033	0.020				0.205	0.125	0.015			- 0.012	0.000					-	
21         0.003         0.003         0.015         0.023         0.115         0.033         0.115         0.033         0.115         0.033         0.115         0.033         0.115         0.033         0.115         0.033         0.115         0.033         0.115         0.035         0.115         0.035         0.115         0.033         0.115         0.035         0.115         0.035         0.115         0.035         0.115         0.035         0.115         0.035         0.015         0.015         0.015         0.015         0.015         0.013         0.015         0.013         0.015         0.003         0.013         0.015         0.013         0.015         0.013         0.013         0.013         0.003         0.	20				0.038	0.018				0.005	0 123				0.015	- 0.80					-	
21.3       International constraints       International constraints       International constraints       International constraints         22       0.005       0.003       0.0095       0.073       0.128       0.003         23       -       0.003       0.125       0.105       -         23       -       0.003       0.125       0.105       -         24       0.003       0.022       0.103       0.003       -         25       -       0.003       0.038       0.165       -         26       -       0.013       0.013       -       -         27       0.013       0.013       0.013       -       -         28       -       0.013       0.013       -       -         30.2       0.028       -       -       -       -         31.2       0.010       0.028       -       -       -         32.2       -       0.010       -       -       -       -         31.2       0.028       -       -       -       -       -       -         32.2       0.020       -       -       -       -       -       -       - <td>20</td> <td></td> <td>0.003</td> <td></td> <td>0.003</td> <td>0.020</td> <td></td> <td></td> <td></td> <td>0.115</td> <td>0.033</td> <td></td> <td></td> <td></td> <td></td> <td>0.155</td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td></td>	20		0.003		0.003	0.020				0.115	0.033					0.155					-	
22       0.005       0.003       0.095       0.073       0.128       0.003         22.2       -       -       -       0.008       -         23.2       -       0.003       0.125       0.003       -         23.2       -       -       -       0.003       -         24       0.003       0.013       0.165       -         25       0.003       0.038       0.165       -         26       -       -       0.013       0.088       -         27       0.013       0.013       0.010       0.003       -         28       0.014       0.028       0.013       0.010       0.003         30       0.0290       -       -       -       -       -         30.2       0.0290       -       -       -       -       -         31.2       0.0120       -       -       -       -       -       -         32.2       0.010       0.0120       -       -       -       -       -       -         32.2       0.028       -       -       -       -       -       -       -       -       -<	21.3				-	-				0.003	-					-					-	
22.2       -       -       -       -       0.008       -         23       -       0.003       0.125       0.105       -         24       0.003       0.052       0.103       0.163       -         25       0.003       0.038       0.165       -       -         26       -       0.013       0.088       -       -         27       0.013       0.013       0.088       -       -         27       0.013       0.013       0.003       0.003       0.003         29       0.143       0.013       -       -       -         30.2       0.028       -       -       -       -       -         31.2       0.003       0.028       -       -       -       -       -         31.2       0.010       -       -       -       -       -       -       -         32.2       0.028       -	22				0.005	0.003				0.095	0.073					0.128					0.003	
23       -       0.003       0.125       0.105       -         23.2       -       -       -       0.003       0.63       -         24       0.003       0.003       0.038       0.165       -         25       0.005       -       -       -       -         25.2       0.005       -       -       -       -         26       -       0.013       0.013       -       -         27       0.013       0.013       -       -       -       -         28       0.143       0.010       0.003       0.003       -       -       -         30       0.218       0.003       - <td< td=""><td>22.2</td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td>-</td><td>-</td><td></td><td></td><td></td><td></td><td>0.008</td><td></td><td></td><td></td><td></td><td>-</td><td></td></td<>	22.2				-					-	-					0.008					-	
23.2       -       -       -       0.003       -       -         24       0.003       0.052       0.103       0.165       -         25       0.003       0.038       0.165       -       -         26       -       -       0.013       0.088       -         27       0.013       0.013       0.013       -       -         28       0.143       0.010       0.003       0.003       -         30       0.218       0.003       -       -       -       -         31       0.045       -       -       -       -       -         31.2       0.028       -       -       -       -       -         32.2       0.010       -       -       -       -       -         34.2       0.005       -       -       -       -       -       -         34.2       0.028       - <td>23</td> <td></td> <td></td> <td></td> <td>-</td> <td></td> <td></td> <td></td> <td></td> <td>0.003</td> <td>0.125</td> <td></td> <td></td> <td></td> <td></td> <td>0.105</td> <td></td> <td></td> <td></td> <td></td> <td>-</td> <td></td>	23				-					0.003	0.125					0.105					-	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	23.2				-					-	-					0.003					-	
25     0.005     0.038     0.05     -       25.2     0.005     -     -     -       26     -     0.013     0.088     -       27     0.013     0.013     -     -       28     0.143     0.010     0.003       30     0.218     0.003     -       30.2     0.028     -     -       31.1     0.045     -     -       32.2     0.010     -     -       34.2     0.005     -     -       34.2     0.005     -     -       34.2     0.055     0.556     0.449     0.813     0.371     0.576     0.277     0.279     0.596     0.713	24				0.003					0.052	0.103					0.163					-	
25.2       0.003       -       -       0.013       - <t< td=""><td>25</td><td></td><td></td><td></td><td></td><td></td><td></td><td>0.005</td><td></td><td>0.003</td><td>0.038</td><td></td><td></td><td></td><td></td><td>0.165</td><td></td><td></td><td></td><td></td><td>-</td><td></td></t<>	25							0.005		0.003	0.038					0.165					-	
20     0.013     0.03     -       27     0.013     0.013     -       28     0.143     0.010     0.003       29     0.218     0.003     -       30     0.290     -     -       31     0.045     -     -       31.2     0.003     -     -       32.2     0.120     -     -       34.2     0.005     -     -       0.120     -     -     -       34.2     0.005     -     -       0.0237     0.028     0.056     0.827     0.735     0.556	25.2							0.005			0.012					0.000					-	
28     0.143     0.003       29     0.218     0.003       30     0.290     -       31     0.045     -       31.2     0.083     -       32.2     0.010     -       34.2     0.005     -       45     -     -	20							0 013			0.015					0.000					-	
29       0.218       0.003         30       0.290       -         30.2       0.028       -         31       0.045       -         31.2       0.083       -         32.2       0.010       -         34.2       0.005       -         45       -       -         9       0.851       0.237       0.028       0.066       0.827       0.735       0.556       0.449       0.813       0.371       0.756       0.981       0.576       0.277       0.279       0.596       0.713	28							0.143								0.010					0.003	
30       0.290       -         30.2       0.028       -         31       0.0045       -         31.2       0.083       -         32       0.010       -         32.2       0.120       -         34.2       0.005       -         45       -       -	29							0.218								0.003						
30.2       0.028       -         31       0.045       -         31.2       0.083       -         32       0.010       -         32.2       0.120       -         34.2       0.005       -         45       0.058       0.056       0.827       0.735       0.556       0.449       0.813       0.371       0.756       0.693       0.777       0.279       0.596       0.713	30							0.290								-						
31       0.045       -         31.2       0.003       -         32.2       0.100       -         34.2       0.005       -         45       -       -	30.2							0.028								-						
31.2     0.083     -       32     0.010     -       32.2     0.120     -       34.2     0.005     -       45     -     -	31							0.045								-						
32     0.010     -       32.2     0.120     -       34.2     0.005     -       45     -     -	31.2							0.083								-						
32.2         0.120         -           34.2         0.005         -           45         -         0.003	32							0.010								-						
3+2         0.003           45         0.003	32.2							0.120								-						
P 0.851 0.237 0.028 0.543 0.058 0.066 0.827 0.735 0.556 0.449 0.813 0.371 0.756 0.633 0.505 0.981 0.576 0.777 0.779 0.596 0.713	34.2							0.005								- 0.003						
1 9.944 9.641 9.949 9.774 9.900 9.774 9.900 9.974 9.133 9.371 9.132 9.371 9.133 9.371 9.101 9.774 9.740 9.711	P	0.851	0.237	0.028	0.543	0.058	0.066	0.827	0.735	0.556	0 449	0.813	0 371	0.756	0.633	0.505	0.981	0.576	0 277	0 279	0.596	0.713

## kit in the Paraguayan population calculated from poppr and pegas in ${\it R}$

P- Hardy-Weinberg exact test based on 3200 shufflings, N- number of individuals typed

*Table 4: Allele frequencies and descriptive statistics of the 15 STR loci included in the Identifiler<sup>®</sup> PCR amplification kit in the* 

Allele	CSF1P0 N: 300	D2S1338 N: 300	D3S1358 N: 300	D5S818 N: 300	D7S820 N: 300	D8S1179 N: 300	D13S317 N: 300	D16S539 N: 300	D18S51 N: 300	D19S433 N: 300	D21S11 N: 300	FGA N: 300	TH01 N: 300	TPOX N: 300	vWA N: 300
6 7 8 9	0.002 0.003 0.022			0.060 0.008 0.042	0.010 0.142 0.107	0.010 0.012	0.078 0.147	0.030 0.160					0.247 0.268 0.102 0.118 0.260	0.413 0.082	
10 11 11.2	0.228 0.295			0.063 0.423	0.252	0.072	0.205	0.125	0.012	0.002 0.008 0.003			0.005	0.057 0.345	0.003
12 12.2	0.370		0.007	0.277	0.198	0.115	0.300	0.258	0.143	0.095				0.103	-
13 13.2 14	0.008		0.013	0.002	0.003	0.320	0.063	0.133	0.237	0.207					0.003
14.2 15 15.2			0.340	0.002		0.108		0.003	0.143	0.020 0.137 0.063					0.087
16 16.2 17		0.033	0.257			0.033			0.123	0.038 0.022 0.003		0.002			0.277
18 19		0.095 0.152	0.100 0.008						0.048	0.000		0.010 0.077			0.172 0.077
20 21		0.127 0.027							0.017 0.010			0.080 0.135			0.003
21.2 22 22.2		0.058							0.005			0.003 0.162 0.005			
23 23.2 24		0.138										0.123 0.005 0.170			
25 25.2 26		0.052									0.005	0.162			
27 28 20											0.020 0.115	0.005			
30 30.2											0.257 0.022				
31 31.2 32											0.037 0.110 0.013				
32.2 33 33.2											0.123 0.002 0.053				
34.2 35											0.010 0.002				
H <sub>exp</sub> H <sub>obs</sub>	0.720	0.872	0.768	0.721	0.801	0.799	0.814	0.806	0.861	0.810	0.837	0.871 0.863	0.776	0.691 0.727	0.771 0.733
P	0.051	0.824	0.479	0.274	0.862	0.182	0.574	0.932	0.415	0.421	0.407	0.164	0.528	0.161	0.563
PD MPE	0.872 0.439	0.968 0.741	0.910 0.904 0.575	0.884	0.925 0.682	0.935 0.931 0.568	0.941 0.545	0.934 0.932 0.605	0.965 0.961 0.715	0.935 0.637	0.951 0.721	0.966 0.721	0.915 0.911 0.533	0.831 0.844 0.471	0.912 0.914 0.482

Identifiler® PCR amplification kit in the Paraguayan population

Hesp - expected heterozygosity, Hebs- observed heterozygosity, P - Hardy-Weinberg exact test based on 3200 shufflings, N- number of individuals typed, PE- Power of Exclusion, PD- Power of Discrimination, MPE- Mean Power of Exclusion

## *Table 5: Allele frequencies and Exact Test for Hardy-Weinberg Equilibrium of the 15 STR loci included in the Identifiler*<sup>®</sup> *PCR*

$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	i Penta D
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	N: 200
5       0.003       0.020         6       -       -       -         7       -       0.115       0.075       0.003       0.005       0.233       0.013         8       0.428       0.070       -       0.033       0.013       -       0.013       0.110       0.023       0.090       0.165         9       0.078       0.150       -       0.015       0.033       0.018       0.130       0.135       0.003       0.133       0.338         9.3       -       -       -       -       -       -       0.218       -       -       0.218       -       -       0.010       0.423         10       0.058       0.050       0.088       0.058       0.233       0.003       0.218       -       -       0.043       0.042       -       0.020       0.003       0.003       0.003       0.004       0.423       0.044       0.044       0.044       -       -       -       -       -<	0.006
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	0.003
1       0.113       0.013       0.003       0.003       0.003       0.013         8       0.428       0.070       0.033       0.013       0.13       0.110       0.023       0.009       0.165         9       0.078       0.150       0.015       0.030       0.018       0.130       0.135       0.003       0.133       0.338         9.3       -       -       -       -       -       -       -       0.013       0.133       0.338         9.3       -       -       -       -       -       -       -       0.013       0.023       0.098       0.018       0.013         10       0.058       0.050       0.048       0.058       0.233       0.003       0.113       0.230       0.003       0.098       0.010       0.423         11       0.330       0.003       0.200       0.008       0.113       0.445       0.248       0.003       0.233       0.005       0.004       0.0423         12       0.108       -       0.280       0.223       0.085       0.158       -       -       -       -       -       -       -       -       -       -       -	-
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	0.003
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	0.015
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	0.220
11       0.330       0.003       0.200       0.008       0.113       0.445       0.242       0.003       0.223       0.025       0.095       0.003       0.001         12       0.108       -       0.280       0.123       0.138       0.215       0.425       -       0.280       0.223       0.085       0.158       -       0.005       0.015       0.020       0.020       0.020       0.020       0.020       0.020       0.020       0.020       0.020       0.020       0.020       0.020       0.021       0.015       0.020       0.020       0.021       0.015       0.021       0.015       0.020       0.021       -       -       -       -       0.020       -       -       -       -       0.021	0 233
12       0.108       -       0.280       0.123       0.138       0.215       0.425       -       0.280       0.223       0.085       0.158       -       -         12.2       -       -       -       -       -       -       -       0.020       0.023       0.085       0.158       -       -       12.2         13       -       0.178       0.090       0.105       0.148       0.065       -       0.013       0.015       0.020       0.080       0.283       0.243       0.001         13.2       -       -       -       -       -       -       -       -       0.048       -         14       0.055       0.070       0.240       0.090       0.013       0.008       -       0.050       0.15       0.020       0.230       0.278       0.021         142       -       -       -       -       -       -       -       -       -       0.040       -         142       -       -       -       -       -       -       -       -       0.040       -         15       0.093       0.003       0.120       0.005       0.003 <td< td=""><td>0.125</td></td<>	0.125
12.2       -       -       -       -       -       -       0.020       -         13       -       0.178       0.090       0.105       0.148       0.065       -       0.013       0.015       0.020       0.080       0.283       0.243       0.001         13.2       -       -       -       -       -       -       -       -       0.013       0.015       0.020       0.080       0.283       0.243       0.001         14       0.055       0.070       0.240       0.090       0.013       0.008       -       0.050       0.015       0.005       0.140       0.230       0.278       0.021         14.2       -       -       -       -       -       -       -       -       0.040       -         15       0.093       0.003       0.123       0.020       0.005       0.003       -       0.015       0.003       0.110       0.093       0.160       0.48	0.180
13         -         0.178         0.090         0.105         0.148         0.065         -         0.013         0.015         0.020         0.080         0.283         0.243         0.043           13.2         -         -         -         -         -         -         -         -         -         0.013         0.015         0.020         0.080         0.283         0.243         0.04           14         0.055         0.070         0.240         0.090         0.013         0.008         -         -         -         -         -         0.013         0.015         0.055         0.140         0.230         0.278         0.001           14.2         -         -         -         -         -         -         -         0.040         -           14.2         -         -         -         -         -         -         -         0.040         -           15         0.093         0.003         0.120         0.005         0.003         -         0.015         0.03         0.110         0.093         0.160         0.48	-
13.2       -       -       -       -       -       -       -       -       0.048       -         14       0.055       0.070       0.240       0.090       0.013       0.008       -       0.050       0.015       0.005       0.140       0.230       0.278       0.00         14.2       -       -       -       -       -       -       0.040       -         15       0.093       0.003       0.120       0.005       0.003       -       0.015       0.003       0.110       0.093       0.160       0.48	0.123
14         0.055         0.070         0.240         0.090         0.013         0.008         -         0.050         0.015         0.005         0.140         0.230         0.278         0.01           14.2         -         -         -         -         -         -         -         -         0.040         -         -         0.040         -         -         -         -         0.040         -         0.040         -         0.44	-
14.2         -         -         -         -         -         -         0.040         -           15         0.093         0.003         0.123         0.120         0.005         0.003         -         0.015         0.003         0.110         0.093         0.160         0.45	0.088
15 0.093 0.003 0.123 0.120 0.005 0.003 - 0.015 0.003 0.110 0.093 0.160 0.44	-
	0.003
	-
	-
10 0.290 0.143 0.00 0.003 0.038 0.233 0.178 0.018 0.023 0.35	-
	-
10.5 0.040 - 0.003	0.005
	0.005
18 0185 0.073 0.020 0.003 0.115 0.103 0.010 0.020 -	
18.3 0.065	
19 0.058 0.033 0.020 0.205 0.123 0.015 - 0.050 -	
19.3 0.003 - 0.013 -	
20 - 0.038 0.018 0.250 0.123 0.080 -	
21 0.003 0.003 0.020 0.115 0.033 0.155 -	
21.3 0.003	
22 0.005 0.003 0.095 0.073 0.128 0.00	
23 - 0.003 0.123 0.103 -	
23.2 0.003	
24 0.00 0.00 0.00	
252 0.005	
26 - 0.013 0.088 -	
27 0.013 0.013	
28 0.143 0.010 0.00	
29 0.218 0.003	
30 0.290 -	
30.2 0.028 -	
31 0.045 -	
31.2 0.083	
32	
24.2 0.120 -	
94.2 0.000	
P 0.851 0.237 0.028 0.543 0.058 0.066 0.827 0.735 0.556 0.449 0.813 0.371 0.756 0.633 0.505 0.981 0.576 0.277 0.279 0.57	0.713

	1.	C• . •	1 • . •	.1	D			1 .•	1	1.1	c		1		· -	•
am	nn	tication	1211 IN	the	Paraa	uavan	nonu	lation	calcul	atod	trom	nonnr	and	neads	in R	<i>.</i>
um	$\nu u$	110011011	n n n n n	inc	Iuiuz	navan	vvvu	anon	cuicui	aica	HOIII	vvvvr	unu	Degus	uu	۰.
							r - r					r • r r		F - ()		

P- Hardy-Weinberg exact test based on 3200 shufflings, N- number of individuals typed

Table 6: Kolmogorov-Smirnov (KS) test results comparing the allele frequency distributions ofshared loci between the two kits to each other for equivalence.

Locus	<i>P</i> -value	D
CSF1P0	1.000	0.100
D2S1338	0.736	0.273
D3S1358	0.975	0.200
D5S818	0.975	0.200
D7S820	1.000	0.100
D8S1179	0.975	0.200
D13S317	0.975	0.200
D16S539	0.675	0.300
D18S51	0.912	0.188
D19S433	0.890	0.200
D21S11	0.589	0.267
FGA	0.997	0.143
TH01	0.975	0.200
TPOX	1.000	0.100
vWA	0.675	0.300

Table 7: Allele frequencies and descriptive statistics of the 15 STR loci included in the Athos® PCR and Identifiler<sup>®</sup> PCR

amplification kits

Allele	CSF1P0	D2S1338	D3S1358	D5S818	D7S820	D8S1179	D13S317	D16S539	D18S51	D19S433	D21S11	FGA	TH01	TPOX	vWA
-	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500	N: 500
5									0.001				0.267		
7	0.002			0.066	0.008								0.267		
8	0.002			0.010	0.129	0.015	0.075	0.023					0.097	0.419	
9	0.021			0.037	0.118	0.008	0.148	0.148					0.124	0.080	
9.3	-			-	-	-	-	-					0.243	-	
10	0.230			0.061	0.243	0.082	0.060	0.120	0.010	0.004			0.007	0.057	
11	0.276			0.432	0.261	0.087	0.203	0.251	0.010	0.006				0.339	0.003
11.2	-			-	-	-	-	-	-	0.002				-	-
12	0.392		0.004	0.252	0.208	0.132	0.292	0.267	0.135	0.095				0.105	-
12.2	-		-	-	-	-	-	-	-	0.014					-
13	0.067		0.013	0.133	0.028	0.305	0.155	0.170	0.091	0.257					0.002
13.2	-		-	-	-	-	-	-	-	0.044					-
14	0.010		0.076	0.006	0.004	0.241	0.000	0.018	0.238	0.285					0.051
14.2			0.358	-	0.001	0 102	0.001	-	0 1 2 5	0.028					-
15.2			0.558	0.005	0.001	0.102	0.001	0.005	0.155	0.063					0.089
16		0.035	0.255			0.027			0 131	0.038					0.282
16.2		-	-			-			-	0.019					-
16.3		-	-			-			0.001	-					-
17		0.213	0.182			0.001			0.121	0.003		0.002			0.324
18		0.103	0.101						0.058			0.010			0.177
18.2		-	-						-			0.001			-
19		0.140	0.011						0.032			0.006			0.069
19.2		-							-			0.001			-
20		0.125							0.025			0.080			0.002
21		0.029							0.007			0.143			0.001
21.2		0.064							0.005			0.003			
22.2		-							-			0.006			
23		0.133							-			0.116			
23.2		-							-			0.004			
24		0.105							0.001			0.170			
25		0.046										0.163			
25.2		-									0.005	-			
26		0.007									0.001	0.071			
27											0.017	0.005			
20											0.126	0.004			
30											0.225	0.001			
30.2											0.024	-			
31											0.040	-			
31.2											0.099	-			
32											0.012	-			
32.2											0.122	-			
33											0.001	-			
33.2											0.049	-			
34.2											0.008	-			
45											0.001	0.001			
He	0.713	0.873	0.758	0.723	0 799	0.807	0.815	0.800	0.862	0.815	0.832	0.874	0.777	0.690	0.770
Ho	0.700	0.884	0.756	0.716	0.830	0.794	0.786	0.798	0.848	0.816	0.852	0.870	0.784	0.714	0.728
Р	0.048	0.923	0.889	0.008	0.848	0.414	0.138	0.975	0.651	0.539	0.779	0.365	0.192	0.384	0.450
PE	0.867	0.970	0.904	0.885	0.929	0.937	0.941	0.930	0.966	0.942	0.951	0.972	0.913	0.851	0.912
PD	0.868	0.969	0.901	0.885	0.926	0.937	0.939	0.930	0.965	0.940	0.951	0.970	0.911	0.846	0.914
MPE	0.428	0.763	0.520	0.453	0.656	0.588	0.573	0.595	0.691	0.629	0.699	0.735	0.570	0.450	0.473

Hesp- expected heterozygosity, Hobs- observed heterozygosity, P - Hardy-Weinberg exact test based on 3200 shufflings, N- number of individuals typed, PE- Power of Exclusion, PD- Power of Discrimination, MPE- Mean Power of Exclusion

#### LIST OF REFERENCES

- 1. *The World Factbook: Paraguay.* Central Intelligence Agency 19 October 2017; Available from: <a href="http://www.cia.gov/library/publications/the-world-factbook/geos/pa.html">www.cia.gov/library/publications/the-world-factbook/geos/pa.html</a>.
- Paraguay 2017 Crime & Safety Report. Paraguay 2017 Crime & Safety Report 13 Mar.
   2017; Available from: www.osac.gov/pages/ContentReportDetails.aspx?cid=21442.
- Andrews, C.A. *The Hardy-Weinberg Principle*. Nature Education Knowledge 2010;
   3(10):65:[Available from: <u>www.nature.com/scitable/knowledge/library/the-hardy-weinberg-principle-13235724</u>.
- 4. Slatkin, M., *Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.* Nat Rev Genet, 2008. **9**(6): p. 477-85.
- Fisher, R.A., *Standard calculations for evaluating a blood-group system*. Heredity (Edinb), 1951. 5(1): p. 95-102.
- 6. Brenner, C. and J. Morris. *Paternity index calculations in single locus hypervariable DNA probes: validation and other studies*. in *Proceedings for the International Symposium on Human Identification*. 1990.
- 7. Lewis, P.O. and D. Zaykin, *Genetic Data Analysis: Computer program for the analysis of allelic data*. 2001
- Team, R.C., *R: A language and environment for statistical computing*. 2013, R
   Foundation for Statistical Computing: Vienna, Austria.
- 9. Section, U.C., *Political Map of Paraguay* Nations Online

- Foxall, P.A., et al., Developmental Validation of a Single-Tube Amplification of the 13 CODIS STR Loci, D2S1338, D19S433, and Amelogenin: The AmpFlSTR® Identifiler® PCR Amplification Kit. Journal of Forensic Sciences, 2004. 49(6): p. 1-13.
- Francischini, C.W., D.R. Sumita, and M.R. Whittle, *Development and use of a single-tube four dye, 21plex direct PCR autosomal STR loci amplification assay for human identification and relationship testing*. Forensic Science International: Genetics
   Supplement Series, 2013. 4(1): p. e264-e265.
- Guo, S.W. and E.A. Thompson, *Performing the exact test of Hardy-Weinberg proportion for multiple alleles*. Biometrics, 1992. 48(2): p. 361-72.
- Butler, J., *Genetics and genomics of core STR loci used in human identity testing*. Journal of Forensic Sciences, 2006. vol. 51(no. 2): p. 253-265.
- Ghasemi, A. and S. Zahediasl, Normality tests for statistical analysis: a guide for nonstatisticians. Int J Endocrinol Metab, 2012. 10(2): p. 486-9.
- 15. Kolmogorov-Smirnov Test.
- 16. Kamvar, Z.N., J.F. Tabima, and N.J. Grunwald, *Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction*. PeerJ, 2014.
   2(281): p. e281.
- 17. Paradis, E., *Pegas*. Pegas: Population and Evolutionary Genetics Analysis System, 2017.
  0.10.
- Armstrong, R.A., *When to use the Bonferroni correction*. Ophthalmic Physiol Opt, 2014.
   34(5): p. 502-8.

- 19. Grattapaglia, D., et al., *Brazilian population database for the 13 STR loci of the AmpFISTR Profiler Plus and Cofiler multiplex kits*. Forensic Science International 2001.
  118: p. 91-94.
- 20. Berardi, G., U. Toscanini, and E. Raimondi, *STR data for PowerPlex 16 System from Buenoes Aires population, Argentina*. Forensic Science International, 2003. 134(2-3): p. 222-224.