DNA Allele Frequencies and other Forensic Parameters of 21 GlobalFiler™ STRs of the El Salvador population

By

JESSICA CASTREJON

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Forensic Science

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Robert Rice, Chair

Ruth Dickover

Henry Erlich

Committee in Charge

2022

Acknowledgements

I would like to express my gratitude to Dr. Cassandra Calloway (Children's Hospital Oakland Research Institute) for the initial project guidance and technical advice, Dr. Henry Erlich (Children's Hospital Oakland Research Institute) for helping me continue my research in his laboratory, Dr. Steven Mack (University of California, San Francisco) for helping and giving me guidance on the statistical analysis of the data, Tanya Tannous and Ava Wu (UC Berkeley) for helping me receiving, processing and punching samples and extracting DNA, Audra McKinzie from ThermoFisher for software and GlobalFiler™ analysis training, the nonprofit organization, Pro-Busqueda for collecting the samples and for trusting us to take on this project, Dr. Ruth Dickover for the mentorship and advice, Dr. Robert Rice for being my thesis chair and helping me finish my thesis and Chris Torres Garcia for guiding and helping me fill out all the paperwork required to graduate.

I would also like to thank my friends for giving me the inspiration to pursue a master's degree and the emotional support. Lastly, I would like to thank my family for their never-ending encouragement and support.

Abstract

An allele frequency database can be used to determine how common or rare an allele or a combination of alleles is. Allele frequencies can be used to determine random match probabilities in forensic cases, paternity testing, mass disasters, cold cases and missing persons investigations. The purpose of this study is to create a population database that represents the 6.7 million people estimated to be living in El Salvador. We determined the allele frequencies of the 21 loci of the GlobalFiler™ PCR Amplification Kit for the El Salvador population. The DNA samples were collected by Pro-Busqueda, a non-profit organization in El Salvador from unrelated volunteers to be representative of the country. We received 762 DNA samples, extracted DNA from 502 samples, quantified, amplified and separated DNA fragments by size and determined the alleles from 360 samples, genotyped and generated DNA profiles for 317 samples. We used STRAF, an online tool to calculate allele frequencies, test loci for Hardy-Weinberg equilibrium and linkage disequilibrium, determine if there is population substructure, as well as calculate relevant forensic parameters such as the power of discrimination (PD or 1-PM), match probability (PM), polymorphism information content (PIC), the power of exclusion (PE) and typical parental index (TPI). We determined the 21 autosomal Globalfiler™ loci used for the database are in Hardy-Weinberg equilibrium and, therefore, can be used for genotyping probabilities. The loci are also in linkage equilibrium, are independent from one another and therefore the product rule can be used to calculate random match probabilities and likelihood ratios. We used GenePop v 4.7.5 to perform a Fisher's Exact G-Test for population differentiation within the country. After validating the allele frequency database, we compared it to the United States Hispanic allele frequency database. We determined that the El Salvador and the United States Hispanic databases were significantly different from one another and cannot be

used interchangeably. The allele frequency database generated from this study will the first to contain allele frequencies from 13 of the 14 departments in the country. This allele database will be a useful tool in helping identify and reunite victims of the country's civil war with family members.

Table of contents

1. Introduction

Allele frequency databases are used to determine how common or rare an allele is or a combination of alleles are. Allele frequencies can be used to calculate random match probabilities, a useful tool in forensic cases, paternity testing, mass disasters, cold cases and missing persons investigations. To create an allele frequency database and generate allele frequencies, samples need to be collected from a representative group of individuals [1].

Once DNA profiles have been generated, the allele frequency database is validated by showing that the distribution of alleles adhere to Hardy-Weinberg equilibrium expectations and the loci are not in linkage disequilibrium. Statistical tests are used to ensure that allele frequencies are reasonable and based on genetic inheritance principles [1]. According to the Hardy-Weinberg principle, in a large population with random mating, no selection, no mutation and no migration, allele frequencies and genotype frequencies are constant from generation to generation [2]. When a population is in Hardy-Weinberg equilibrium, genotype frequencies can be estimated from allele frequencies. Linkage disequilibrium is the non-random association of alleles at different loci. When loci are in linkage equilibrium, then the product rule can be applied to allele frequencies to calculate random match probabilities and other metrics.

Allele frequencies can be used to calculate random match probabilities, paternity indexes and various important forensic parameters such as the power of discrimination (PD or 1-PM), match probability (PM), polymorphism information content (PIC), the power of exclusion (PE) and typical parental index (TPI). Random match probability is the probability of a specific profile occurring in a specific population based on the observed allele frequencies within that population [1]. Match probability (PM) is the probability of a match between two unrelated individuals [4].

Power of discrimination (PD) is the potential power of a set of loci to differentiate between two people chosen at random in a population [1].

Paternity index is a likelihood ratio that compares how much more likely it would be to see observed shared alleles if the alleged father is the true father versus a random unrelated man from the same population [1]. Power of exclusion (PE) is the ability of a genetic test to exclude a random individual from the population given the observed alleles of the child and mother [3].

Polymorphism information content (PIC) is a way of determining how informative a locus is in the ability to detect a polymorphism among individuals of a population [3]. PIC can also be interpreted as the probability that the maternal and paternal alleles of a child are deducible or the probability of being able to deduce which allele a parent has transmitted to the child [4]. The greater the PIC value, the more informative the locus is. Typical paternity index (TPI) is the mean paternity index for random non excluded men.

During the years 1980 through 1992 civil war broke out in El Salvador. About 75,000 people died, 2,598 were documented missing persons and an estimated 10,000 were undocumented missing persons [5]. An allele frequency database for the El Salvador population would be useful in identifying missing persons of the civil war and helping to reunite them with their family members. Since 2003, the nonprofit organization Pro-Busqueda and the University of California Berkeley, Human Rights Center, have been collaborating on the Building Justice DNA Reunification Project to help identify and reunite missing people with their families in El Salvador. They investigate reported cases of missing and kidnapped children during the civil war, track down family members, and collect DNA samples from victims and potential blood relatives. They have created a private missing person and family members database. One of their goals is to preserve biological information by collecting and genotyping DNA samples from

older blood relatives in cases where potential victims have not yet been identified. As time goes on more complex kinship calculations will have to be made in order to determine the relativeness relatedness of people.

To reunite identified missing persons of the civil war and their family members, kinship analysis must be performed for the missing persons and potential family members. The STR allele frequencies used by Pro-Busqueda for parental testing and kinship analysis are limited to only 15 STRs, do not have the statistical power to associate more distant relatives, and no longer meet the guidelines for the publication of genetic population data as published in Forensic Science International: Genetics [6, 7, 8, 9, 10, 11].

Previous studies looked at a limited number of STR markers, had a small sample size, took samples from one location or from unknown locations, or did not sample from most or all departments, which are not representative of the whole country [7, 8, 9. 10, 17]. In 2004, a study looked at 6 STRs (HUMTPOX, HUMTH01, HUMVWA, D18S535, D1S1656 and D12S391) in 120 unrelated individuals born and living in San Salvador [8]. In 2005, 17 Y-STRs loci were characterized from 120 unrelated males born and living in San Salvador [10]. In 2006, a study looked at 15 STR loci from the PowerPlex kit from Promega in 296 unrelated individuals, but the collection location was not mentioned [7]. In 2015 a study looked at the 15 STR loci from the AmpFlSTR®Identifiler® PCR Amplification Kit in 109 unrelated Salvadorians living in Spain [19]. In March 2022, a collaborator of this study genotyped and analyzed 58 STRs (27 autosomal, 24 Y-STRs and 7 X-STRs) and 94 autosomal SNPs in Illumina ForenSeq™ Primer Mix A in a sample of 248 men and 143 women from El Salvador [18]. In April 2022, a study looked at 21 STRs for the Salvadoran population using GlobalFiler Express and GlobalFiler STR Amplifcation Kits. They took samples from 683 unrelated individuals from 3 main regions of El

Salvador, central (San Salvador, La Libertad, Chalatenango, Cuscatlan, San Vicente, Cabañas, and La Paz), east (Usulutan, San Miguel, La Union, and Morazan), and west (Ahuachapan, Sonsonate, and Santa Ana).

Establishing a population database for the entire El Salvador population with the 21 STR (short tandem repeat) loci of GlobalFiler™ PCR Amplification Kit would help in cases where the paternity index is below 100, or in more complicated cases where a first-degree common ancestor like a parent or sibling is no longer available for testing. A stronger statistical association in extended family testing can be seen either by collecting data from more genetic markers in the DNA samples being compared or by increasing the pool of reference samples [1].

The purpose of this study is to create a population database and estimate allelic frequencies that represent the approximately 6.7 million people living in El Salvador today and to help identify and reunite victims with family members.

2. Hypotheses

$H_1(1)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are not in Hardy-Weinberg equilibrium.

$H_0(1)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are in Hardy-Weinberg equilibrium.

$H_1(2)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are not in Linkage disequilibrium.

$H_0(2)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are in Linkage equilibrium.

$H_1(3)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population show such statistically significant differences that the allele frequencies are not considered similar enough to be amalgamated into one population database.

$H_0(3)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are similar enough to be amalgamated into one population database.

$H_1(4)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in each department show statistically significant difference, and there appear to be population substructures within some departments.

$H_0(4)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in each department are similar enough to be amalgamated into one population database.

$H_1(5)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are statistically significant different from the allele frequencies of the

Hispanic population in the USA, such that the allele frequencies are not similar enough for the two databases to be amalgamated into one population database.

$H_0(5)$ The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are similar enough to the allele frequencies of the Hispanic population in the USA for the two databases to be amalgamated into one population database.

3. Materials and methods

Sample Collection:

DNA samples were collected by the non-profit organization Pro-Busqueda from volunteers across the 14 departments of El Salvador in 2018. Volunteers were given a consent form and received instructions on how to collect their own DNA. Buccal samples were collected using Bode Buccal DNA Collectors (Bode Technology, Lorton, Virginia) and EasiCollect Plus buccal swabs (Qiagen, Hilden, Germany). After collection was complete samples were shipped to the Calloway Lab at the Children's Hospital Oakland Research Institute in Oakland, California. We received 726 samples from the Calloway lab with only sex and location information.

Figure 1. shows a map of El Salvador with the 14 departments. Ahuachapan (AHU), Cabanas (CB), Chalatenango (CHA), Cuscatlan (CUS), La Libertad (LL), La Paz (LP), La Union (LU), Morazan (MOR), San Miguel (SM), San Salvador (SS), San Vicente (SV), Santa Ana (SA), Sonsonate (SON) and Usulutan (USU).

https://en.wikipedia.org/wiki/Departments_of_El_Salvadorhttps://upload.wikimedia.org/wikipedia/commons/1/10/Departments_of_El_Salvador_named.svg

DNA Sample Extraction:

To extract DNA first we punched round disks from the buccal collectors using 1.2 mm and 3.0 mm Whatman Uni-Core Punchers (GE Healthcare Life Sciences, Chicago, Illinois). We extracted DNA from 502 samples from all 14 departments. We took six 1.2 mm punches or three 3.0 mm punches from the buccal collector FTA cards or Bode collection paper and placed the punches into 1.5 mL microcentrifuge tubes. The punchers were cleaned with bleach, ethanol and placed under UV light for 10 minutes between uses. The DNA was extracted and purified utilizing the QIAamp DNA Investigator kit (Qiagen, Hilden, Germany), manually or automatically using the robot QIAcube (Qiagen, Hilden, Germany), with a final elution of 80 uL TE^{-4} (pH 8.0, 0.1 M EDTA) if extracted manually and 60 uL TE^{-4} (pH 8.0, 0.1 M EDTA) if extracted automatically. We used the manufacturer's protocol for the manual and automatic extraction of FTA and Guthrie cards.

DNA Quantification:

To quantify the DNA, we used the instrument Applied Biosystems 7900 HT Fast Real-Time PCR System with 384-Well Block Module (Applied Biosystems, Bedford, Massachusetts). Due to financial constraints, we only quantified 360 samples out of the 502 that were punched. We did a duplex real time qPCR assay for the quantification of nuclear and mitochondrial DNA [13]. To prepare the samples for DNA quantification, 4 uL of extracted sample, 6 uL of an in-house

mt-nu probe blend and 10 uL of TaqMan Universal PCR Master Mix (Applied Biosystems, Bedford, Massachusetts) were pipetted into a MicroAMP Optical 384-Well Reaction Plate (Applied Biosystems, Bedford, Massachusetts). For quality control, we included a positive control, negative control and two reagent blanks into each plate. For this project we only looked at nuclear DNA.

Oligonucleotide	Sequence (5' → 3')	Location
nuTH01-F	5 ' - agggtatctgggctctgg - 3'	nu RB1 gene
nuTH01-R	5' - ggctgaaaagctcccgattat-3'	nu RB1 gene
mtND1-F	5'- ccctaaaacccgccacatct-3'	mtDNA rCRS 3484-3504
mtND1-R	5'- gagcgatggtgagagctaaggt-3'	mtDNA rCRS 3532-3553
nuTH01 probe	5'- FAM- attcccattggcctgttcctccctt-BHQ 3'	nu RB1 gene
mtND1 probe	5'- VIC- ccatcaccctctacatc- MGB-NFQ 3'	mtDNA rCRS 3506- 3522

Table 1: Oligonucleotide sequences used for the duplex assay.

DNA Amplification:

DNA was amplified using GlobalFiler™ PCR Amplification Kit according to the manufacturer's protocol (Applied Biosystems, Bedford, Massachusetts). Samples were loaded on to a MicroAMP Optical 96-Well Reaction Plate manually with a multi-channel pipette. Final volume was 25 uL (7.5 uL of master mix, 2.5 uL Primer set and 15 uL of DNA). The DNA was amplified in a GeneAmp PCR System 9700 with a gold-plated block (Applied Biosystems, Bedford, Massachusetts) in Max ramping mode for 29 cycles. The initial incubation step time was 1 minute at 95°C, denaturation time was 10 seconds at 94°C, anneal and extension time was 90 seconds at 59°C these steps were repeated for 29 cycles, the final extension time was 10

minutes at 60℃, and the final hold was 4℃ for up to 24 hours. We used the manufacturer's protocol for GlobalFiler™ PCR Amplification kit.

Capillary electrophoresis of DNA fragments:

After DNA amplification, 1 uL of PCR product, 0.4 uL of GeneScan 600 LIZ Size Standard v2.0 (Applied Biosystems, Bedford, Massachusetts) and 9.6 uL of Hi-Di Formamide were added into wells of a MicroAMP Optical 96-Well Reaction Plate. We then sealed the plates and handed them over to the UC Berkeley DNA Sequencing Facility. The facility then removed the seal, added a rubber septum, vortexed and centrifuged the plates, then denatured the DNA in a thermal cycler at 95℃ for 3 minutes. The plates were snap cooled on a cold gold-plated block for 3 minutes. The plates were loaded on to a 3730xl DNA Analyzer (Applied Biosystems, Bedford, Massachusetts) and ran through capillary electrophoresis for DNA separation and sizing. For quality control one positive and one negative control were included in each plate.

DNA Analysis:

To analyze the raw data produced by the genetic analyzer we used Microsatellite Analysis CE Fragment Sizing, a free online program from ThermoFisher. We reached out to ThermoFisher, and they agreed to giving me a tutorial on how to use the program. For analysis of the DNA profiles, we used the CLA GlobalFiler™ PCR Amplification Kit analysis settings that were recommended by ThermoFisher.

Statistical Analysis:

We used STRAF (STR Analysis for Forensics), a free online tool to calculate allele frequencies, test for Hardy-Weinberg equilibrium and linkage disequilibrium, as well as calculate relevant forensic parameters such as the match probability (PM), power of discrimination (PD), polymorphism information content (PIC), power of exclusion (PE) and typical paternity index (TPI).

STRAF was also used to determine if there is population substructure using F-statistics. GenePop v 4.7.5 was used to perform a Fisher's exact G-Test for population differentiation. We also used a Fisher's exact G-Test to determine if the El Salvador and the US Hispanic allele frequency databases were different.

4. Results

Pro-Busqueda collected726 samples from the 14 departments in the El Salvador population. Samples were sent to the Calloway Lab at the Children's Hospital Oakland Research Institute in Oakland, California. We punched disks from 502 samples and placed them into tubes then extracted and quantified the DNA. We amplified and sequenced 360 samples, after analysis we generated 317 complete DNA profiles. From these DNA profiles we were able to determine allele frequencies of 21 loci in the GlobalFiler™ PCR Amplification Kit for 13 of the 14 departments in El Salvador. We then validated the allele frequency database by determining the loci are in Hardy-Weinberg equilibrium and linkage equilibrium.

Allele Frequencies:

We used the online tool STRAF to calculate allele frequencies for the 21 loci using GlobalFiler™ PCR Amplification Kit. Allele frequencies were calculated using the counting method. Table 2 lists the allele frequencies of the 21 GlobalFiler™ loci. Figure 1 shows a plot for each locus and its allele frequencies. We did not include Amelogenin, the sex determining STR locus, and Y INDEL and DYS391 which are male only STR loci.

Hardy-Weinberg Equilibrium:

To test whether the population is in Hardy-Weinberg Equilibrium (HWE), we used STRAF to do a Fisher's Exact Test with 1000 Monte Carlo permutations to compute the P-value for each locus [4]. At 19 of the 21 loci, the P-value > 0.05. The null hypothesis, which is that the loci are in

Hardy-Weinberg Equilibrium (HWE), is not rejected for these 19 loci. The loci with P-values < 0.05 are D13S317 (P-value=0.0056), D22S1045 (P-value=0.03), and SE33 (P-value=0.0348). We applied the Bonferroni correction, and the new threshold of significance was (adjusted P-value) = 0.00238. The Bonferroni correction is a multiple-comparison correction used when several dependent or independent statistical tests are being performed simultaneously [20]. After the correction was applied, all loci P-values fell above the adjusted threshold of significance. Therefore, the null hypothesis is not rejected. None of the loci in the database deviate from Hardy-Weinberg Equilibrium (HWE) and, therefore can be used for estimating genotype frequencies from allele frequencies. Table 3 shows .lues, as well as observed heterozygosity (Hobs), genetic diversity or expected heterozygosity (GD) for the 21 autosomal GlobalFiler™ loci tested.

Linkage disequilibrium:

To test for linkage disequilibrium, we used STRAF to do a T2 test to compute and give estimates of linkage. A low P-value would indicate there is linkage disequilibrium. All the P-values that were computed for the population were above the adjusted threshold of significance $\alpha = 0.00238$. Therefore, the null hypothesis, which is that the loci are in linkage equilibrium, is accepted. The 21 autosomal GlobalFiler™ loci are not linked, are independent, and randomly recombine into new genotypes. Table 4 shows the pairwise combinations and P-values.

F-Statistics:

F_{st} value per locus is calculated if there is suspected substructure. F_{st} measures the amount of genetic variance that can be explained by population structure based on Wright's F-statistics [15]. An F_{st} value of 0 indicates no differentiation between the subpopulations while a value of 1 indicates complete differentiation [14]. Table 5 shows the computed value per locus for the effect of subpopulations compared to the total population (F_{st}), and the inbreeding coefficient related to the subpopulation (F_{is}) per locus. All the tested GlobalFiler™ loci had small or negative values. We determined there is no population substructure amongst the loci.

We also calculated F_{st} values in a pairwise comparison of subpopulations, to determine whether there is substructure among the 13 departments. We treated the departments as individual populations and compared them to each other. Table 6 shows the F_{st} values in a pairwise comparison of subpopulations. The F_{st} and $F_{is,}$ values are low or negative, meaning that there is no significant difference amongst the 13 departments. The departments can be combined into one database the represents El Salvador.

Fisher's exact G-test:

Before assuming there is linkage or there is no linkage STRAF authors recommend you do another linkage test, we chose to use GenePop v4.75 to test for population diffraction. A Fisher's exact G-test, with 10,000 Monte Carlo permutations, was applied to each department's Globalfiler™ loci data and compared to one another. Table 7 shows the P-values across all loci and the P-value for the entire data set. All P-values per locus have a P-value > 0.05 and across all loci the P-value= 0.275701. With a P-value > 0.05, we accept the null hypothesis which indicates

that there is no significant difference in allele frequencies of these loci between the departments of El Salvador.

The El Salvador allele frequency database was compared to the allele frequency US Hispanic population database that used the AmpFLSTR™ Identifiler™ PCR Amplification Kit which can be found in STRbase. A Fisher's exact G- test was performed using GenePop v 4.75. The P-values calculated for the Globalfiler™ loci were all $P < 0.05$ or negative numbers. Across all loci, the P-values were < 2.21E-30. Table 8 shows P-values for the comparison. Since $P < 0.05$, we reject the null hypothesis which assumes that there is no difference among the populations and accept the alternative hypothesis that the two population databases are significantly different and cannot be used interchangeably or amalgamated.

Other forensic parameters:

We also used STRAF to calculate other important forensic parameters: genetic diversity or expected heterozygosity (GD), observed heterozygosity (Hobs), how many different alleles were seen at each locus (Nall), the power of discrimination (PD or 1-PM), match probability (PM), polymorphism information content (PIC), the power of exclusion (PE) and typical parental index (TPI). The values for each parameter can be seen in Table 9.

The locus with the highest power of discrimination (PD) was SE33 at PD= 0.988765; it also had the highest allele diversity (Nall) and the highest observed heterozygosity (Hobs). The observed heterozygosity value (Hobs=0.930599) was close to the expected heterozygosity value (GD=0.937721). The locus with the lowest power of discrimination (PD) was D22S1045 (PD= 0.838360). D22S1045 showed a high degree of heterozygosity, expected heterozygosity (GD)

ranged from 0.663726 - 0.937721 and observed heterozygosity (Hobs) ranged from 0.643533 - 0.930599. The combined power discrimination was determined to be greater than 99.999999% and the match probability (PM) was 4.3935×10^{-25}. A high combined power of discrimination indicates the ability to differentiate between two people chosen at random in the population with a high degree of certainty.

The polymorphism information content (PIC) values ranged from 0.613941 - 0.932633. A PIC value greater than 0.50 is considered very informative [3]. The typical parental index (TPI) for all the loci was above 1. The combined paternity index (CPI) was 80887762.96 and the probability of paternity ((CPI/CPI+1) X 100) was determined to be 99.9999988%. This means that the database can be useful for the determination of paternity.

Power of exclusion (PE) is the ability of a genetic test to exclude a random individual from the population given the observed alleles of the child and mother. The power of exclusion (PE) values ranged from 0.346405 to 0.858252. When the 21 autosomal GlobalFiler™ loci are combined, the PE is 99.999762%.

5. Discussion

For this study we created an allele frequency database for the 21 autosomal Globalfiler™ loci in the El Salvador population. We validated the database by determining that the 21 loci are in Hardy-Weinberg equilibrium and linkage disequilibrium. To calculate random match probabilities and use the product rule, the loci must be in Hardy-Weinberg equilibrium and linkage equilibrium.

Hardy-Weinberg equilibrium:

We used the Fisher's Exact Test to determine whether the population was in Hardy-Weinberg equilibrium. We saw that three of the loci (D13S317, D21S11 and D8S1179) had a P-value less than 0.05. We used the Bonferroni correction to adjust the threshold of significance to $\alpha = 0.00238$. After the correction was applied, all P-values fell above the threshold of significance. Therefore, we reject the null hypothesis. The 21 autosomal Globalfiler™ loci in the database are in Hardy-Weinberg Equilibrium and can be used for genotyping probabilities.

Linkage disequilibrium:

When testing for linkage all P-values were above the threshold of significance $\alpha = 0.00238$. The null hypothesis is accepted. The allele frequencies generated by the GlobalFiler™ PCR Amplification Kit in the El Salvador population are in linkage equilibrium. The Globalfiler™ loci are independent from one another, therefore the product rule can be used to calculate random match probabilities and likelihood ratios.

F-Statistics:

We tested for population substructure by doing two F-statistics test. First, we tested the loci in the dataset. The F_{st} value per locus ranged from a negative value of -0.0064 to 0.01 and the inbreeding coefficient F_{is} values ranged from -0.0375 to 0.0621. Then we did a pairwise comparison for each department. The F_{st} values ranged from -0.00611 to 0.0127. F_{st} measures the amount of genetic variance that can be explained by population structure. An F_{st} value of 0 indicates no differentiation between the subpopulations while a value of 1 indicates complete differentiation [14]. All the loci had small or negative values. We determined there is no substructure amongst the departments and it is appropriate to combine them into one database.

Fisher's exact G-test:

To test for goodness-of-fit (comparing observed to theoretical expectations) and independence (comparing one database to another) we used the Fisher's exact G-test. The Fisher's exact G-test test gives approximately the same results as the chi-square test [16]. What makes the Fisher's exact G-test the preferred method is that G-values are additive and can be used for more elaborate statistical designs [16]. GenePop v4.7.5 uses the Fisher's exact G-test as default to test population differentiation.

A Fisher's exact G-test was also applied to each department's Globalfiler™ loci and compared to one another. All P-values per locus had a P-value > 0.05; across all loci the P-value= 0.275701. With a P-value > 0.05 we accept the null hypothesis which indicates that there is no significant difference between the departments of El Salvador. A Fisher's exact G- test was also performed on the El Salvador and the US Hispanic population. The P-values were above $P < 0.05$ or had a negative value. Across all loci the P-value $< 2.21 \times 10^{-30}$. Since $P < 0.5$ we reject the null hypothesis which assumes that there is no difference among the populations and accept the

alternative hypothesis that the two databases are significantly different and cannot be used interchangeably or amalgamated.

Other forensic parameters:

We calculated important forensic parameters: the power of discrimination (PD or 1-PM), match probability (PM), polymorphism information content (PIC), the power of exclusion (PE) and typical parental index (TPI).

The combined PD (1-PM) was greater than 99.999999% and the match probability (PM) was 4.3935×10^{-25}. A high combined power of discrimination indicates the ability to differentiate between two people chosen at random in the population with a high degree of certainty. The polymorphism information content (PIC) values were in the range of 0.613941 - 0.932633. A PIC value greater than 0.50 is considered very informative [3]. The combined paternity index (CPI) was 80887762.96 and the probability of paternity ((CPI/CPI+1) X 100) was 99.9999988%. This means that all the loci are useful for the determination of paternity. The power of exclusion (PE) values ranged from 0.346405 to 0.858252. When the 21 autosomal GlobalFiler™ loci are combined, the PE is 99.999762%. Power of exclusion (PE) is the ability of a genetic test to exclude a random individual from the population given the observed alleles of the child and mother. This allelic frequency database is adequate to be used in paternity cases for this population.

Due to financial constraints, we were not able to re-sequence DNA that failed to generate a profile when sequenced, sequence DNA samples from the department Cuscatlán and were not able to genotype the 500 samples needed to publish an allele frequency dataset in Forensic

Science International: Genetics [6]. We were originally given 726 samples and determined allele frequencies from 317 DNA profiles. Reasons for failing to genotype a sample included no DNA, low peak heights, potential trial alleles, migration issues and contamination. Future studies should aim to complete the required 500 samples for an allele frequency publication and to include Cuscatlán, the department not included in this study.

The allele frequencies generated from this study can be used to calculate random match probabilities for the population of El Salvador. The allele frequency database could be combined with Cornejo-Moreno, B.A., et al. allele frequency database [17]. They looked at the 21 STR (short tandem repeat) loci of GlobalFiler™ PCR Amplification Kit in the El Salvador population. Our study was able to sequence alleles that did not show up in in the Cornejo-Moreno, B.A., et al. study and they also had alleles that did not show up in our study. A combined database could generate more accurate allele frequencies for the population and increase random match probabilities. The random match probability is a useful tool in forensic cases, paternity testing, in mass disasters, cold cases and missing persons investigations.

Regarding the Bonferoni correction, despite the widespread use of the Bonferroni method, there has been continuing controversy regarding its use, there are those who believe correction should never be made and those who consider correction should be mandatory [21]. If we were to not use the Bonfferoni correction than the GlobalFiler™ loci would not be in Hardy Weinberg equilibrium. A departure from Hardy Weinberg Equilibrium (HWE) might occur due to a variety of causes such as purifying selection (the selective removal of alleles that are deleterious), inbreeding, population substructure, undetected admixture, copy number variation and genotyping error [22].

6. Conclusion

For this project we developed an allele frequency database for 21 of the Globalfiler™ loci for the El Salvador population. We determined allele frequencies and validated the database by demonstrating that the Globalfiler™ loci are in Hardy-Weinberg equilibrium and linkage equilibrium. We also determined that the alleles generated from each of the 13 departments are not significantly different and can be used in one allele frequency database for the El Salvador population. We do not expect the untested department to be significantly different from the other departments. Additionally, we determined that the El Salvador population is significantly different from the US Hispanic population. The established allele frequency database that was created for this project will be a useful tool for calculations such as the Random Match Probability and Paternity Index for the El Salvador. We hope it will help identify and reunite victims of the country's civil war with family members.

Tables and figures

Table 2: Allele frequencies for the GlobalFiler™ loci in the El Salvador population.

Locus	D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D22S1045	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
n=	634	634	634	634	634	634	634	634	634	634	634	634	634	634	634	634	634	634	634	634	634
5																0.0016					
5.3											0.0032										
6				0.0016	0.0095						0.4069				0.0016						
6.3																0.0016					
7				0.0126	0.0032						0.2697		0.0047	0.0442		0.0158					
8			0.0110	0.0047	0.4953	0.0032			0.0016		0.0852			0.0095	0.0662	0.0836					
9			0.1356	0.0142	0.0394	0.0063		0.0016			0.0883			0.0568	0.2050	0.0410		0.0016			
9.3											0.1420										
10			0.2114	0.2476	0.0426	0.0584		0.0110	0.4164	0.0016	0.0047		0.0063	0.0442	0.1136	0.2713		0.0016	0.0016		
10.3																0.0079					
11			0.2965	0.2823	0.2571	0.0426		0.0047	0.2729	0.0079			0.0568	0.4732	0.2003	0.2729	0.0016	0.0079	0.0268		
11.2										0.0047							0.0016				
11.3									0.0284												
12			0.2445	0.3628	0.1498	0.1309		0.1057	0.0410	0.0536			0.0079	0.2681	0.2066	0.2492	0.0047	0.0268	0.0489		
12.1					0.0032																
12.2										0.0205							0.0047				
13	0.0032	0.0047	0.0994	0.0694		0.2839		0.1404	0.0126	0.2098			0.0095	0.0931	0.1136	0.0521	0.0079	0.2729	0.1451	0.0016	
13.2								0.0016		0.1199							0.0032				
13.3									0.0016												
14	0.0820	0.0710	0.0016	0.0047		0.2934		0.1861	0.2003	0.2839			0.0300	0.0110	0.0915	0.0032	0.0331	0.3785	0.1136	0.0079	
14.2										0.0315											
15	0.4606	0.0694				0.1325		0.1372	0.0237	0.1719			0.4148		0.0016		0.0284	0.2224	0.1672	0.0268	
15.2										0.0536							0.0016				
15.3																			0.0221		
16	0.2524	0.3233				0.0457		0.1025	0.0016	0.0284			0.3738				0.0662	0.0599	0.1909	0.0300	0.0110
16.2										0.0110											
16.3																			0.0536		
17	0.1088	0.3423				0.0032		0.1625					0.0883				0.1041	0.0252	0.0552	0.0741	0.1230
17.2										0.0016							0.0047				
17.3																			0.1309	0.0047	
18	0.0899	0.1372						0.0694				0.0016	0.0063				0.1136	0.0032	0.0016	0.2256	0.0631
18.2												0.0032									
18.3																			0.0315	0.0126	
19	0.0032	0.0442						0.0300				0.1073	0.0016				0.0883		0.0016	0.2208	0.1767
19.1																				0.0016	
19.2												0.0016					0.0016				
19.3																			0.0095	0.0110	
20		0.0079						0.0221				0.0836					0.0710			0.2082	0.1388
21								0.0095				0.1246					0.0142			0.0662	0.0237
21.2												0.0016					0.0047				
22								0.0047				0.0915					0.0110			0.0379	0.0804
22.2												0.0032					0.0237				
23								0.0063				0.1404					0.0016			0.0379	0.2177
23.2												0.0016					0.0268				
24								0.0032				0.1782								0.0221	0.1104
24.2																	0.0237				

25								0.0016				0.1356								0.0063	0.0473
25.2																	0.0457				
26												0.0868								0.0047	0.0016
26.2																	0.0568				
27							0.0079					0.0268									0.0063
27.2																	0.0710				
28							0.0883					0.0110									
28.2																	0.0599				
29							0.2208					0.0016									
29.2							0.0016										0.0568				
30							0.2618														
30.2							0.0237										0.0284				
31							0.0773														
31.2							0.0962										0.0189				
31.3							0.0016														
32							0.0095														
32.2							0.1404										0.0142				
33							0.0016										0.0032				
33.1							0.0016														
33.2							0.0489														
34																	0.0032				
34.1							0.0016														
34.2							0.0110														
35							0.0047														
35.2							0.0016														

Figure 1. Allele frequencies for the GlobalFiler™ loci in the El Salvador population. Plots for each locus were generated with STRAF.

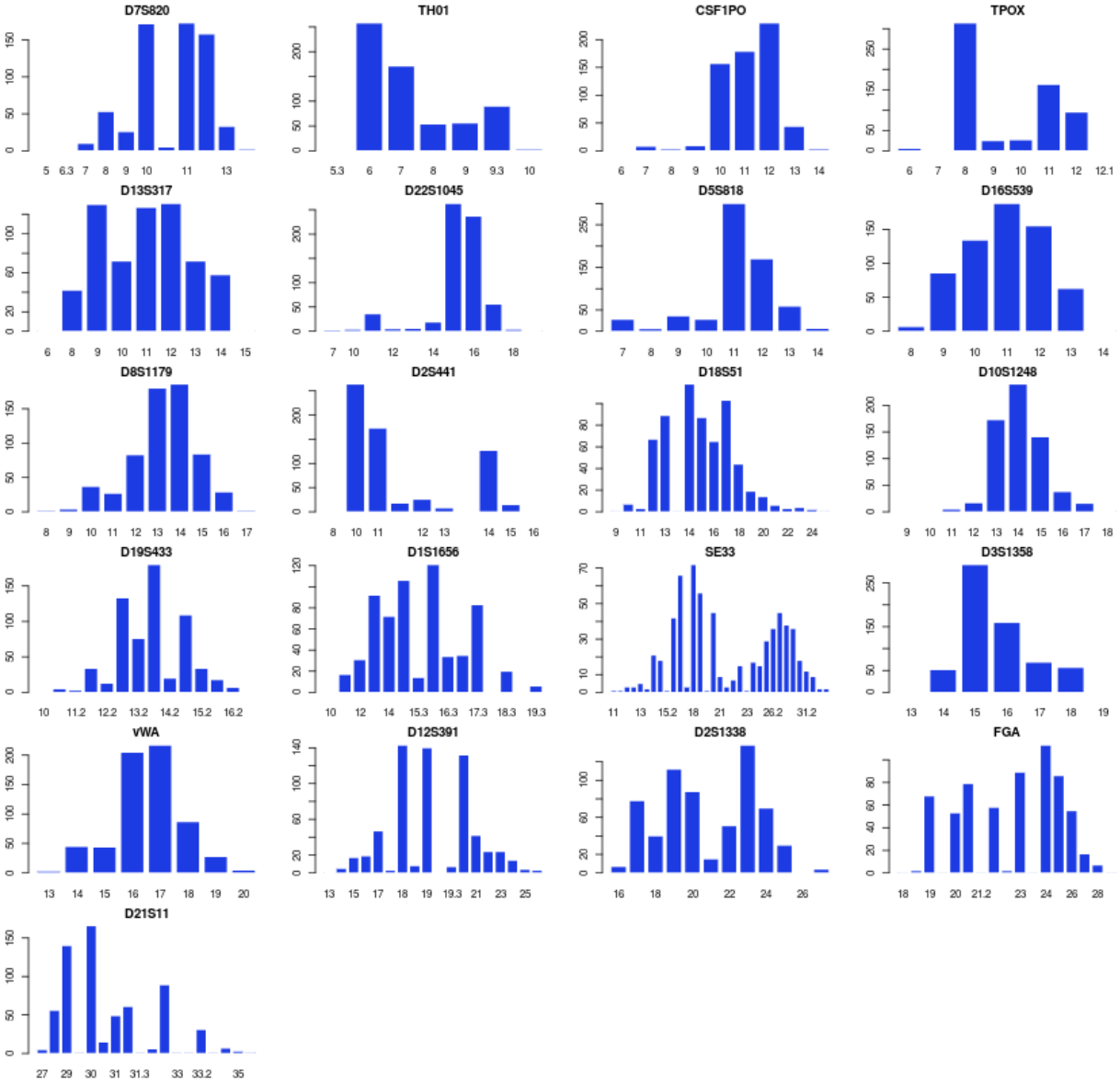

Table 3. Fisher Exact Test computed P-values for the 21 GlobalFiler™ loci tested for Hardy-Weinberg Equilibrium. (P<0.00238). Number of alleles (N), number of different alleles seen within a locus (Nall), observed heterozygosity (Hobs), genetic diversity or expected heterozygosity (GD)

Locus	**N**	**Nall**	**GD**	**Hobs**	**P value**
CSF1PO	634	9	0.7233	0.7192	0.8969
D10S1248	634	10	0.7289	0.7382	0.8329
D12S391	634	18	0.8431	0.8297	0.6264
D13S317	634	9	0.8379	0.8328	0.0056
D16S539	634	7	0.7805	0.7886	0.8294
D18S51	634	18	0.8736	0.8612	0.9022
D19S433	634	14	0.8246	0.8139	0.1296
D1S1656	634	15	0.8753	0.877	0.4276
D21S11	634	18	0.838	0.8107	0.29
D22S1045	634	11	0.6771	0.6593	0.03
D2S1338	634	12	0.8628	0.8454	0.1406
D2S441	634	10	0.7099	0.6656	0.3277
D3S1358	634	7	0.6986	0.7098	0.4034
D5S818	634	8	0.6893	0.6435	0.171
D7S820	634	11	0.7794	0.7886	0.2436
D8S1179	634	10	0.7925	0.7697	0.6472
FGA	634	17	0.8807	0.8801	0.9442
SE33	634	33	0.9377	0.9306	0.0348
TH01	634	7	0.7276	0.7066	0.6442
TPOX	634	8	0.6637	0.6435	0.5041
vWA	634	8	0.7488	0.7129	0.1068

Table 4. Pair wise combinations tested for Linkage Disequilibrium and P-values for the 21 GlobalFiler™ loci.

	vWA	TPOX	TH01	SE33	FGA	D8S1179	D7S820	D5S818	D3S1358	D2S441	D2S1338	D22S1045	D21S11	D1S1656	D19S433	D18S51	D16S539	D13S317	D12S391	D10S1248	CSF1PO
CSF1PO	0.9548	0.954764	0.417092	0.662112	0.535882	0.885346	0.767078	0.691146	0.61862	0.095704	0.697908	0.031976	0.300724	0.58681	0.152334	0.120636	0.195762	0.129928	0.750676	0.048896	
D10S1248	0.688514	0.463466	0.190828	0.599926	0.66569	0.432122	0.482228	0.350188	0.96184	0.55087	0.349882	0.039308	0.01737	0.718912	0.399608	0.925902	0.789214	0.44575	0.100558		0.048896
D12S391	0.420862	0.10523	0.604406	0.447318	0.022206	0.383402	0.687882	0.076166	0.455396	0.442468	0.324934	0.745158	0.784336	0.022854	0.964916	0.658728	0.504738	0.28418		0.100558	0.750676
D13S317	0.407606	0.090452	0.282688	0.968412	0.771292	0.702418	0.87695	0.892136	0.244872	0.093004	0.922366	0.601126	0.872128	0.194506	0.933854	0.980818	0.703534		0.28418	0.44575	0.129928
D16S539	0.526276	0.204716	0.621696	0.210572	0.157884	0.89886	0.86563	0.830894	0.355782	0.380992	0.048456	0.61124	0.533488	0.599236	0.544522	0.978798		0.703534	0.504738	0.789214	0.195762
D18S51	0.219804	0.142986	0.876484	0.008264	0.713542	0.97012	0.144282	0.89813	0.441234	0.984636	0.719916	0.438866	0.01591	0.172526	0.15504		0.978798	0.980818	0.658728	0.925902	0.120636
D19S433	0.38691	0.17719	0.087042	0.990556	0.502718	0.267306	0.801776	0.87818	0.33024	0.443976	0.243514	0.083048	0.455696	0.31159		0.15504	0.544522	0.933854	0.964916	0.399608	0.152334
D1S1656	0.998702	0.688728	0.790332	1	0.6795	0.112308	0.357452	0.843132	0.167538	0.16936	0.443544	0.628808	0.614618		0.31159	0.172526	0.599236	0.194506	0.022854	0.718912	0.58681
D21S11	0.33239	0.23819	0.26852	0.91276	0.640028	0.683512	0.926834	0.117792	0.1497	0.546998	0.028728	0.041506		0.614618	0.455696	0.01591	0.533488	0.872128	0.784336	0.01737	0.300724
D22S1045	0.90536	0.029214	0.254892	0.00347	0.281084	0.832274	0.229322	0.458878	0.249858	0.62635	0.739544		0.041506	0.628808	0.083048	0.438866	0.61124	0.601126	0.745158	0.039308	0.031976
D2S1338	0.21218	0.10633	0.489486	0.00889	0.122898	0.390408	0.105554	0.278182	0.004546	0.97325		0.739544	0.028728	0.443544	0.243514	0.719916	0.048456	0.922366	0.324934	0.349882	0.697908
D2S441	0.152926	0.833856	0.01903	0.448974	0.37534	0.03768	0.928444	0.649202	0.096242		0.97325	0.62635	0.546998	0.16936	0.443976	0.984636	0.380992	0.093004	0.442468	0.55087	0.095704
D3S1358	0.433924	0.704766	0.980714	0.262314	0.202998	0.485706	0.454466	0.094814		0.096242	0.004546	0.249858	0.1497	0.167538	0.33024	0.441234	0.355782	0.244872	0.455396	0.96184	0.61862
D5S818	0.90952	0.561744	0.770638	0.87505	0.382454	0.68288	0.223292		0.094814	0.649202	0.278182	0.458878	0.117792	0.843132	0.87818	0.89813	0.830894	0.892136	0.076166	0.350188	0.691146
D7S820	0.8486	0.251972	0.330426	0.48	1	0.441192		0.223292	0.454466	0.928444	0.105554	0.229322	0.926834	0.357452	0.801776	0.144282	0.86563	0.87695	0.687882	0.482228	0.767078
D8S1179	0.776052	0.534004	0.051124	0.217224	0.18755		0.441192	0.68288	0.485706	0.03768	0.390408	0.832274	0.683512	0.112308	0.267306	0.97012	0.89886	0.702418	0.383402	0.432122	0.885346
FGA	0.880922	0.224314	0.261354	0.027636		0.18755	1	0.382454	0.202998	0.37534	0.122898	0.281084	0.640028	0.6795	0.502718	0.713542	0.157884	0.771292	0.022206	0.66569	0.535882
SE33	0.592812	0.26006	0.483014		0.027636	0.217224	0.48	0.87505	0.262314	0.448974	0.00889	0.00347	0.91276	1	0.990556	0.008264	0.210572	0.968412	0.447318	0.599926	0.662112
TH01	0.493274	0.272398		0.483014	0.261354	0.051124	0.330426	0.770638	0.980714	0.01903	0.489486	0.254892	0.26852	0.790332	0.087042	0.876484	0.621696	0.282688	0.604406	0.190828	0.417092
TPOX	0.083428		0.272398	0.26006	0.224314	0.534004	0.251972	0.561744	0.704766	0.833856	0.10633	0.029214	0.23819	0.688728	0.17719	0.142986	0.204716	0.090452	0.10523	0.463466	0.954764
vWA		0.083428	0.493274	0.592812	0.880922	0.776052	0.8486	0.90952	0.433924	0.152926	0.21218	0.90536	0.33239	0.998702	0.38691	0.219804	0.526276	0.407606	0.420862	0.688514	0.9548

Table 5. F statistics P-values for the 21 GlobalFiler™ loci. * Number of alleles (N), number of different alleles seen within a locus (Nall), observed heterozygosity (Hobs), genetic diversity or expected heterozygosity (GD).

Locus	N	Nall	GD	Hobs	Fst	Ht	Fis	P_value
CSF1PO	634	9	0.7233	0.7192	0.01	0.7222	-0.007	0.956136
D10S1248	634	10	0.7289	0.7382	-0.0014	0.7267	-0.0049	0.952902
D12S391	634	18	0.8431	0.8297	-0.001	0.8416	0.0073	0.840369
D13S317	634	9	0.8379	0.8328	-0.0039	0.8383	0.0128	0.571362
D16S539	634	7	0.7805	0.7886	0.0043	0.7804	-0.0261	0.859435
D18S51	634	18	0.8736	0.8612	0.0054	0.8754	0.0077	0.782904
D19S433	634	14	0.8246	0.8139	-0.002	0.8241	0.0038	0.225262
D1S1656	634	15	0.8753	0.877	-0.0018	0.8764	0.0017	0.853006
D21S11	634	18	0.838	0.8107	-0.0028	0.8406	0.0511	0.04384
D22S1045	634	11	0.6771	0.6593	0.0019	0.6802	0.0206	0.172618
D2S1338	634	12	0.8628	0.8454	-0.0023	0.8658	0.0204	0.384257
D2S441	634	10	0.7099	0.6656	0.0037	0.7132	0.0588	0.409927
D3S1358	634	7	0.6986	0.7098	0.007	0.6931	-0.0068	0.785617
D5S818	634	8	0.6893	0.6435	0.0052	0.6785	0.0621	0.149838
D7S820	634	11	0.7794	0.7886	0.0007	0.7785	-0.0375	0.069917
D8S1179	634	10	0.7925	0.7697	0.0012	0.7943	0.038	0.006409
FGA	634	17	0.8807	0.8801	-0.0015	0.877	0.001	0.802554
SE33	634	33	0.9377	0.9306	-0.0052	0.9367	0.0127	0.730219
TH01	634	7	0.7276	0.7066	-0.0051	0.7262	0.025	0.70228
TPOX	634	8	0.6637	0.6435	-0.0064	0.6667	0.0181	0.40841
vWA	634	8	0.7488	0.7129	-0.0017	0.7504	0.0465	0.193922

Table 6. F_{st} values in a pairwise comparison of subpopulations. Ahuachapan (AHU), Cabanas (CB), Chalatenango (CHA), Cuscatlan (CUS), La Libertad (LL), La Paz (LP), La Union (LU), Morazan (MOR), San Miguel (SM), San Salvador (SS), San Vicente (SV), Santa Ana (SA), Sonsonate (SON) and Usulutan (USU).

	AHU	**CB**	**CHA**	**LL**	**LP**	**LU**	**MOR**	**SA**	**SM**	**SON**	**SS**	**SV**
USU	-0.00664	0.000687	0.005097	0.000929	0.00535	-0.00175	-0.00234	-0.00156	0.000523	0.000809	-0.0019	-0.00435
SV	-0.00995	-0.00611	-0.0011	-0.00056	0.002058	-0.00274	-0.00188	-0.00489	-0.00327	-0.00273	-0.0039	
SS	-0.00459	-0.00166	0.001926	-0.00018	0.004712	0.002815	0.000409	-0.0027	0.000691	-0.00503		
SON	0.000495	-0.00288	-0.00212	-0.00242	-0.006	0.000938	-0.00214	-0.00041	0.000528			
SM	-0.00773	9.58E-05	0.004096	0.004874	0.012719	-3.7E-05	0.002824	0.001595				
SA	-0.00871	-0.00099	-0.0036	-0.00097	0.003878	0.00509	-0.00141					
MOR	-0.00194	0.001032	0.00233	0.000932	-0.00041	-0.00248						
LU	-0.00416	0.00133	0.003309	0.005273	0.003219							
LP	0.001978	0.004823	0.003219	0.007251								
LL	-0.00136	0.001977	0.00756									
CHA	-8.6E-06	0.001163										
CB	-0.00788											
AHU												

Table 7. Fisher's exact G-Test. P-value across all loci for the El Salvador population. P value=0.05 and threshold of significance α= 0.00238.

Locus	**P-value**
D3S1358	0.125572
vWA	0.89221
D16S539	0.121786
CSF1PO	0.295542
TPOX	0.187858
D8S1179	0.19452
D21S11	0.095468
D18S51	0.115218
D2S441	0.77408
D19S433	0.170996
TH01	0.705392
FGA	0.283644
D22S1045	0.093796
D5S818	0.28929
D13S317	0.795464
D7S820	0.90567
SE33	0.996336
D10S1248	0.366732
D1S1656	0.48625
D12S391	0.499622
D2S1338	0.842862

Table 8. Fisher's exact G-Test. P-value, across 15 loci for the El Salvador population vs US Hispanic population. P-value=0.05 and threshold of significance $\alpha = 0.00238$.

Locus	**P-value**
D3S1358	0.00003
vWA	0
D16S539	0.00003
CSF1PO	0.58569
TPOX	0.01053
D8S1179	0.26657
D21S11	0.06977
D18S51	0.19954
D19S433	0.00009
TH01	0
FGA	0.00142
D5S818	0.04043
D13S317	0.02454
D7S820	0.00101
D2S1338	0

Table 9. Other forensic parameters. genetic diversity or expected heterozygosity (GD), observed heterozygosity (Hobs), how many alleles were seen at each locus (Nall), polymorphism information content (PIC), match probability (PM), the power of discrimination (PD), the power of exclusion (PE) and typical parental index (TPI).

Locus	N	Nall	GD	PIC	PM	PD	Hobs	PE	TPI	P_value
CSF1PO	634	9	0.72327458	0.67238222	0.12627253	0.87372747	0.7192429	0.45865357	1.78089888	0.8969
D10S1248	634	10	0.72890597	0.68217854	0.12300849	0.87699151	0.73817035	0.48974649	1.90963855	0.8329
D12S391	634	18	0.84305869	0.82357927	0.04238275	0.95761725	0.829653	0.65518135	2.93518519	0.6264
D13S317	634	9	0.83790074	0.81549531	0.05217486	0.94782514	0.83280757	0.66127628	2.99056604	0.0056
D16S539	634	7	0.78046058	0.74422578	0.08487496	0.91512504	0.78864353	0.5781355	2.36567164	0.8294
D18S51	634	18	0.87361271	0.85880979	0.03067997	0.96932003	0.86119874	0.7170524	3.60227273	0.9022
D19S433	634	14	0.82457478	0.80157466	0.05645394	0.94354606	0.81388013	0.62505032	2.68644068	0.1296
D1S1656	634	15	0.87529714	0.86105613	0.03255083	0.96744917	0.87697161	0.74866195	4.06410256	0.4276
D21S11	634	18	0.83800539	0.81766546	0.0457264	0.9542736	0.81072555	0.61909595	2.64166667	0.29
D22S1045	634	11	0.67706729	0.62023395	0.16163958	0.83836042	0.65930599	0.36815401	1.46759259	0.03
D2S1338	634	12	0.86276855	0.84623044	0.03766581	0.96233419	0.84542587	0.68586935	3.23469388	0.1406
D2S441	634	10	0.70992869	0.66123949	0.12790455	0.87209545	0.66561514	0.37709686	1.49528302	0.3277
D3S1358	634	7	0.69861607	0.65531963	0.13536805	0.86463195	0.70977918	0.44355038	1.72282609	0.4034
D5S818	634	8	0.68928691	0.64637618	0.1386918	0.8613082	0.64353312	0.34640488	1.40265487	0.171
D7S820	634	11	0.77935922	0.74378091	0.08993024	0.91006976	0.78864353	0.5781355	2.36567164	0.2436
D8S1179	634	10	0.79251075	0.762215	0.07438625	0.92561375	0.76971609	0.54409589	2.17123288	0.6472
FGA	634	17	0.88072919	0.8670667	0.02872951	0.97127049	0.88012618	0.75502853	4.17105263	0.9442

SE33	634	33	0.93772083	0.93263328	0.01123506	0.98876494	0.93059937	0.85825191	7.20454545	0.0348
TH01	634	7	0.72756041	0.68479801	0.11204211	0.88795789	0.70662461	0.43858269	1.70430108	0.6442
TPOX	634	8	0.66372638	0.61394077	0.16036581	0.83963419	0.64353312	0.34640488	1.40265487	0.5041
vWA	634	8	0.74876035	0.70892512	0.10306601	0.89693399	0.71293375	0.44855151	1.74175824	0.1068

References

1. Butler, J. M. (2010). Chapter 11 - Statistical Interpretation: Evaluating the Strength of Forensic DNA Evidence. *Fundamentals of Forensic DNA Typing. J. M. Butler. San Diego, Academic Press*: 229-258.
2. Gou. S. W., and Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics,* 46, 361-371.
3. Serrote, C. M. L., Reiniger, L. R. S., Silva, K. B., Rabaiolli, S. M. D. S, and Steanel, C. M. (2020). Determining the Polymorphism Information Content of a molecular marker. *Gene.* 726, 144175.
4. Gouy, A. and Zieger, M., (2017). STRAF—A convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics*, 30, 148-151.
5. Harvey-Blankenship, M., Pham, P. N., Shigekane, R., (2010). Genetic Tracing, Disappeared Children and Justice*, Innocenti Working Papers no*, 2010-12.
6. Gusmao, L., Butler, J. M., Linacre, A., Parson, W., Roewer, L., Schneider, P. M., & Carracedo, A. (2017). Revised guidelines for the publication of genetic population data. *Forensic Science International: Genetics,* 30, 160-163.
7. Monterrosa, J. C., Morales, J. A. and García, O. (2006). Genetic Variation for 15 Short Tandem Repeat Loci in an El Salvadoran (Central America) Population. *Journal of Forensic Sciences*, 51, 451–452.
8. Martínez-Jarreta, B., Vásquez, P., Abecia, E., Garde, M., de Blas, I., Budowle, B. (2004). Autosomic STR loci (HUMTPOX, HUMTH01, HUMVWA, D18S535, D1S1656 and D12S391) in San Salvador (El Salvador Central America). *Journal of Forensic Sciences*, 49, 633–634.

9. Morales, J., Monterrosa, J., Alvarez, J., Entrala, C., Lorente, J., Lorente, M., Budowle, B., and Villanueva, E. (2002). Population Data on Nine STR Loci in an El Salvadoran (Central American) Sample Population. *Journal of Forensic Sciences*, 47(4), 1-2.
10. Martínez-Jarreta, B., Vásquez, P., Abecia, E., Budowle, B., Luna, A., and Peiró, F. (2005). Characterization of 17 Y-STR Loci in a Population from El Salvador (San Salvador, Central America) and Their Potential for DNA Profiling. *Journal of Forensic Sciences*, 50(5), 1243-6.
11. Engels WR. (2009) Exact Tests for Hardy–Weinberg Proportions. *Genetics,* 183(4), 1431-1441.
12. Butler, J.M. (2015). Chapter 14 - Relationship Testing: Kinship Statistics. *Advanced Topics in Forensic DNA Typing: Interpretation. J. M. Butler. San Diego, Academic Press*: 349-401.
13. Timken, Mark D., Swango, Katie L., Orrego, Cristian, & Buoncristiani, Martin R. (2005) A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples. *Journal of Forensic Sciences*, 50(5), 1044-1060.
14. Bird, K. A., An, H., Gazave, E., Gore, M. A., Pires, J. C., Robertson, L. D., et al. (2017). Population structure and phylogenetic relationships in a diverse panel of Brassica rapa L. *Frontier Plant Science*, 8:321. https://doi.org/10.3389/fpls.2017.00321.
15. Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, Vol 19, 395–420. doi: 10.1111/j.1558-5646.1965.tb01731.x

16. McDonald, J.H. 2014. *Handbook of Biological Statistics (3rd ed.)*. *Sparky House Publishing, Baltimore, Maryland.* 68-76.

17. Cornejo-Moreno, B. A., Henríquez-García, M. A., Ramos-Aleman, S. L., Romero, X. P., Lazo-Mena, E., de Pleitez, M., Vullo, C. M., Ge, J., & Budowle, B. (2022). Population data for 21 STRs for the Salvadoran population using GlobalFiler Express and GlobalFiler STR Amplification Kits. *International journal of legal medicine*, *136*(2), 565–567. https://doi.org/10.1007/s00414-021-02709-6

18. Casals, F., Rasal, R., Anglada, R., Tormo, M., Bonet, N., Rivas, N., Vásquez, P., & Calafell, F. (2022). A forensic population database in El Salvador: 58 STRs and 94 SNPs. *Forensic science international. Genetics*, *57*, 102646. https://doi.org/10.1016/j.fsigen.2021.102646

19. Muñoz, P., Pinto de Erazo, E.L., Baeza, C. et al. Genetic polymorphism of 15 STR loci in El Salvador. *Int J Legal Med* 129, 991–993 (2015). https://doi.org/10.1007/s00414-015-1148-8

20. Weisstein, Eric W. "Bonferroni Correction." From MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/BonferroniCorrection.html

21. Armstrong R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131

22. Chen, B., Cole, J. W., & Grond-Ginsbach, C. (2017). Departure from Hardy Weinberg Equilibrium and Genotyping Error. *Frontiers in genetics*, *8*, 167. https://doi.org/10.3389/fgene.2017.00167