

Research Article

Open Access

Action Protocols in DNA Identification of Isolated Populations

Maria Saiz MS^{1#}, Maria J Alvarez-Cubero^{1#}, Juan C Alvarez^{1,2}, Luis J Martinez-Gonzalez^{2*} and Jose A Lorente^{1,2}

¹Laboratory of Genetic Identification, Dept of Legal Medicine, Universidad de Granada, Granada, Spain

²GENyO Centre for Genomics and Oncological Research: Pfizer - Universidad de Granada - Junta de Andalucía, Spain

[#]Authors contributed equally to this work

Abstract

This paper presents the basic problems and difficulties that can be found when working with isolated populations as far as propose some approaches in the analysis of this type of populations. 13 autosomal STRs have been analyzed and statistical forensic parameters, such as observed heterozygosity and power of discrimination have been determined, in samples from isolated and non-isolated populations. Samples were amplified with AmpFISTR® Identifier kit (Applied Biosystems) and PowerPlex 16® kit (Promega). For DNA typing, an ABI PRISM 310 Genetic Analyzer was used, and the analysis was performed with GeneMapper ID-X Software v1.1. PowerStats software and SPSS v15.0 were used to calculate forensic and other parameters. The analysis is based on the comparison of three main population groups (large, medium, and small), and an estimation of the forensic parameters, such as discrimination power (PD), observed heterozygosity (Ho), and combined PD, was obtained. The obtained results reveal that heterozygosity and PD are lower in aboriginal populations than in other populations. This research contributes to defining the decrease in allele presence caused by the lower size of a population as far as endogamy processes. To obtain an efficient human identification, it is necessary to generate independent databases for all the different populations, including small isolated groups. Action protocols to these kinds of populations have to be adapted: type as many markers as possible, not only autosomal STR but also mitochondrial DNA and sexual chromosome markers and characterize those one that better describe the population, and also get genetic information of close relatives as far as direct references of the individual.

Keywords: Discrimination power; Identification; Isolated populations; Observed heterozygosity; STRs

Introduction

The most common use of STRs is in forensic cases [1]. However, STR population data has also been used for admixture estimation and structure inference in different populations [2,3]. When a sample is identified with a specific genetic profile, the probability of discrimination can differ depending on the laboratory where it was processed and on the STR markers chosen and their population frequency. Heterozygosity and homozygosity are both parameters that show the polymorphism and efficacy of a genetic marker.

Even today, some parts of the world are still isolated due to geographic, linguistic, or cultural factors. Endogamy is characteristic of isolated groups which creates a problem in forensic casework. Loss of heterozygosity and the presence of same haplotypes are found at a higher proportion in these populations because of aboriginal identity preservation. Most relationships occur among the community members, and this phenomenon of endogamy usually continues within the next generations. Furthermore, there is a strong tradition of endogamy and a preference for consanguineous unions. Although there are a lot of similarities in terms of anthropology, geographic ethnicity, or culture, the populations may still exhibit major genetic differentiation [4]. When working with forensic samples coming from isolated and small populations, some problems in genetic identification can be found even when a direct relative is available.

Materials and Methods

Subjects and samples

Three kinds of populations were used in this essay: large populations living in a vast territory, medium or large populations living in small territory, and small isolated populations.

This project was carried out using the genetic data of 13 autosomal short tandem repeats (STRs) polymorphisms in three classes of populations (small isolated populations, medium populations, and large populations). The samples were analyzed with different kits; hence, the project was reduced to 13 CODIS markers (D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX, and FGA). The small isolated populations are Chakma, Čakovec, Shiite Muslim, OtomiIxmiquilpan, Huastecos, AdiPanggi, AdiKomkar, Kaqchiquel, Kiche, Mam, and Qeqchi. The medium populations are compounded by Liaoning, Kuala Lumpur, Dobruja, Northern Poland, Karamoja, Rio de Janeiro, East Timor, Granada, and Ovambo. The big populations are Macedonia, Morocco, Somalia, Guatemalan Mestizo, Spain, Mexico-Chihuahua, Paraguay, Costa Rica, Brazil, and Jalisco. Some of these populations, such as Granada, Guatemalan Mestizo, Spain, Paraguay, Ecuador, Costa Rica, Brazil, Mexico-Chihuahua, Kaqchiquel, Kiche, Mam, and Qeqchi, were analyzed and processed in the Laboratory of Genetic Identification (University of Granada, Spain). The others were collected from previously published frequency data (For more information about the populations see supplementary data Table S1).

***Corresponding author:** Luis J Martinez-Gonzalez, GENyO. Centre for Genomics and Oncological Research: Pfizer - Universidad de Granada - Junta de Andalucía, Avd. Ilustración 114 Granada, Granada 18016, Spain, Tel: +34958715501; Fax: +34 958 637 071; E-mail: luisjavier.martinez@genyo.es

Received December 30, 2013; **Accepted** February 07, 2014; **Published** February 14, 2014

Citation: Maria Saiz MS, Alvarez-Cubero MJ, Alvarez JC, Martinez-Gonzalez LJ, Lorente JA (2014) Action Protocols in DNA Identification of Isolated Populations. J Forensic Res 5: 218 doi:10.4172/2157-7145.1000218

Copyright: © 2014 Maria Saiz MS, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Sample collection followed the recommendations of GEP-ISFG, and all the samples were collected under the informed consent of the donors. Each single project used in this article has been approved by the Ethical Committee.

Populations

Special focus is given to isolated populations, which are all characterized by cultural, linguistic, and/or geographical factors (Figure 1).

Some population groups, like Kaqchiquel, Kiche, Mam, Qeqchi, and Huastecos, are Mayan tribes. There are more than twenty Mayan languages and a huge variety of populations groups that explains the high grade of isolation present in these Mayan groups [5]. Most of these groups belong to the same language family, Mayan Quichean-Mamean, although they preserve their native language [6]. K'iche', Kaqchikel and Q'eqchi' are into the Greater Quichean language group while Mam language is situated in the Greater Mamean group and Huastecos, another Mexican tribe, call their language "Teenek" [7].

The other isolated populations also have their own characteristics that help to preserve their aboriginal identity. Chakmas, an ethnic group from Bangladesh, speaks Changma [8]; OtomiIxmiquilpan, a Mexican tribe, continues to use their own language, Hñahñu [7]; and AdiPanggi and AdiKomkar are dialects of Adi, a language spoken in India, China, and Bhutan [9]. Although they abandoned some traditional customs and trades, the Romani ethnic minority Čakovec kept a distinct archaic dialect of the Romanian (*Limba d' bjaš*) as a predominant language [10]. Shiite Muslims from Tamilnadu, India, besides speaking the local languages, have their own language called *Lisānu l-Dā'wat*, "The language of the Dā'wat". All of these populations also have their own culture and customs, and profess the same religion. Geographical isolation is also present in these and other similar groups. While Mayan tribes reside in the highlands of Guatemala, Chakmas live near the foothills of the Himalayas; AdiPanggi and AdiKomkar are part from little populations of Tibet, and OtomiIxmiquilpan inhabit a semi-deserted area at an altitude of 1700-1800 m.

Besides linguistic, cultural, and geographical studies, genetic analyses have also confirmed that aboriginal and isolated populations

are closely related and are more similar in terms of population diversity [11].

DNA analysis

The samples were amplified with AmpFlSTR®Identifiler® kit (Applied Biosystems, Foster City, CA) [12] and a GeneAmp System 2400 thermal cycler. Some populations, like Paraguay, Costa Rica, and Brazil, were amplified with PowerPlex 16®kit (Promega) [13]. A total of 13 markers were chosen for the statistical analysis (D3S1358, TH01, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, CSF1PO, vWA, D8S1179, TPOX, and FGA). Genotyping of nuclear STRs was achieved using an ABI PRISM® 310 Genetic Analyzer (Applied Biosystems). Fragment sizes were analyzed by GeneScan v3.1 [14] and GeneMapper® ID-X Software v1.1 (Applied Biosystems) [15].

Statistic analysis

Allele frequencies were determined by direct counting. Gene diversity and haplotype diversity were calculated according to Nei [16,17]. All the forensic statistical parameters were recalculated with the PowerStats software [18]. Other statistic parameters, such as power of discrimination (PD; medium, maximum, and minimum), observed heterozygosity (Ho; maximum and minimum), and combined PD (-log 10 (1-combined PD)), were calculated by SPSS v15.0.

Results

In order to compare the three classes of populations, the forensic parameters were estimated. High values of PD were obtained in large (0.915 ± 0.004) and medium populations (0.916 ± 0.007) in contrast to small isolated populations (0.847 ± 0.055). Moreover, higher Ho values were found in big (0.783 ± 0.007) and medium (0.786 ± 0.013) populations compared to those in small isolated populations (0.7662 ± 0.030). These results were not as different as those in PD, but isolated populations still had the lowest values. The differences in PD and Ho between small isolated populations, and medium and large populations are significant (Figure 2), more details in Table 2.

The values of the minimum and maximum Ho achieved the biggest differences in isolated populations, whereas in large and medium populations, these values had lesser differences between them

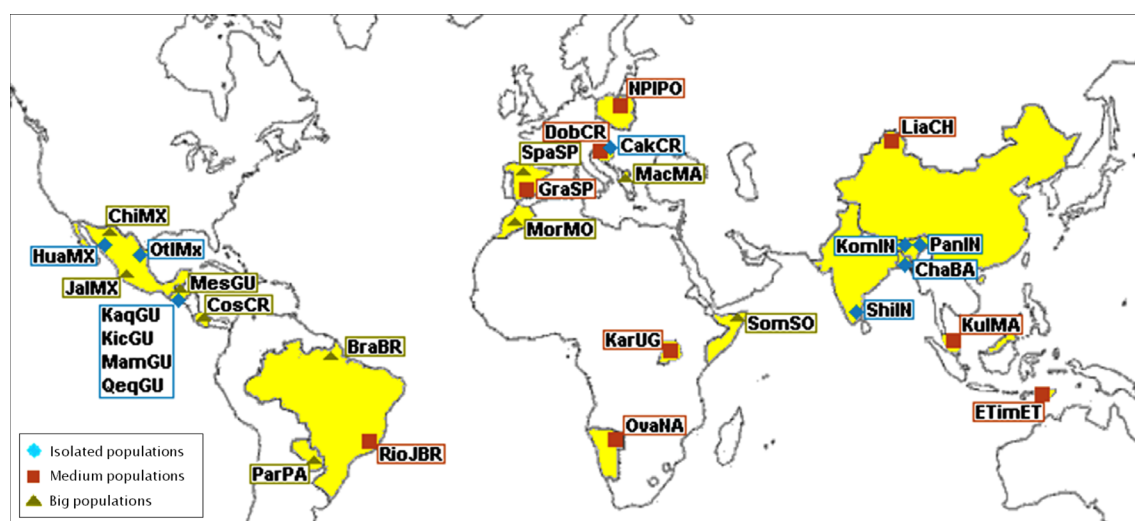


Figure 1: Isolated Populations which were characterized by cultural, linguistic, and/or geographical factors.

(Figure 3). Similar results have been obtained in the analysis of PD. Studies about these types of populations have revealed an increase in homozygosity and a decrease in genotypic variability. Thus, the power of discrimination, typical paternity index, and power of exclusion are lower, too (data not shown).

If we present combined PD, the differences can be also noticed. As

in previous analyses and representations, the lowest values are found in isolated populations, while the highest values are achieved in medium and big populations (Figure 4).

Furthermore, the gene diversity is also different among these populations. If PD is taken into account, whereas TH01 has the less discrimination capacity and FGA the highest in isolated populations,

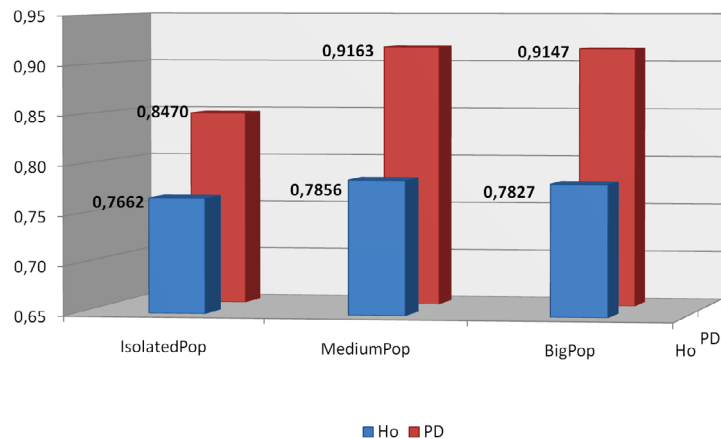


Figure 2: The differences in PD and Ho between small isolated populations, and medium and large populations.

Sample Info	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH0	D13S31	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
Case 1 Mother	12,13	29,29	12,12	9,11	14,15	6,6	9,12	9,11	19,22	13,13	16,17	8,11	14,14	9,12	23,25
Case 1 Minor	12,14	29,31.2	12,12	11,12	14,15	6,9	9,9	9,11	22,23	13,15	16,17	8,11	14,15	11,12	24,25
Case 1 Father	11,12	30,31	12,12	11,12	15,18	9,9	9,11	9,13	22,23	13,13.2	13,17	11,12	14,16	12,13	22,25
Case 2 Mother	14,15	29,31.2	9,10	7,10	15,15	6,7	8,11	11,12	19,23	13,14	16,18	8,12	12,16	11,12	18.2,25
Case 2 Minor	13,14	29,31.2	10,11	10,12	15,15	6,6	11,12	11,12	22,23	13,15.2	15,16	12,12	16,21	11,12	18.2,23
Case 2 Father	12,13	29,33.2	11,11	9,10	15,16	6,9.3	11,12	10,12	20,23	13,13.2	16,18	11,12	15,18	11,11	21,23

In italics and bold, markers that exclude the father when studying minor and father. Only in bold, those markers that exclude the father when studying mother, minor and father

Table 1: Genetic profiles of mother, minor and father in two paternity cases.

Sample Info	PD p-value	OR(95%CI)	Ho p-value	OR(95%CI)
Isolated vs. large populations	0.0001*	-0.0676(-0.0839- -0.0513)	0.0753	-
Isolated vs. medium populations	0.001*	-0.0693(-0.0859 - -0.0526)	0.0474*	-0.0194(-0.0385- -0.0002)
Medium vs. large populations	0.0737	-	0.7303	-

*This difference is considered statistically significant.
vs: Versus
OR: Odds Ratio
CI: Confidence Interval

Table 2: PD and Ho statistical analysis.

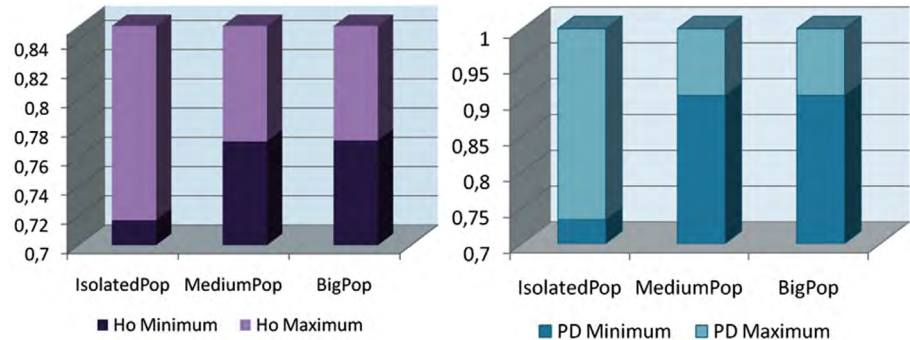


Figure 3: The values of the minimum and maximum Ho achieved the biggest differences in isolated populations.

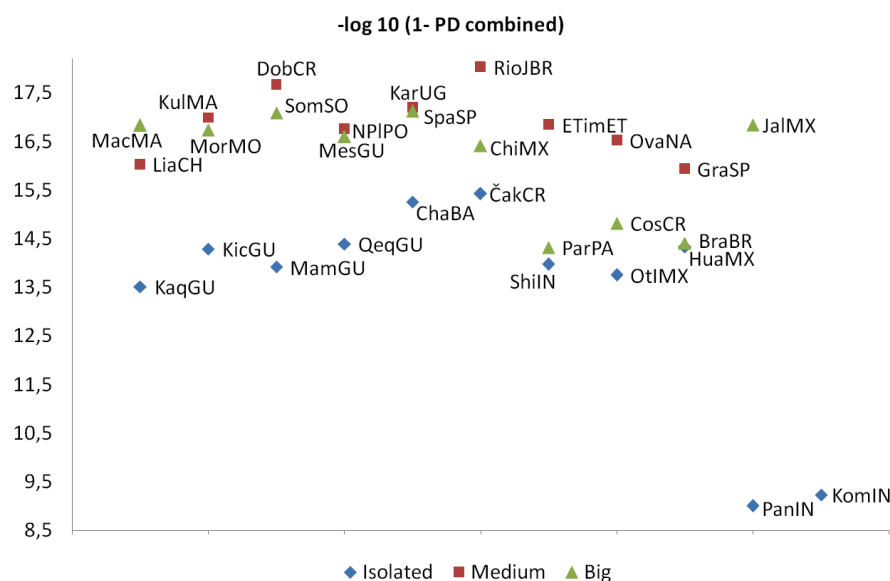


Figure 4: Graphical analysis of Discrimination power (PD).

in medium populations are D5S818 and D18S51 respectively; and TPOX and D18S51 in large size populations. When studying H_o , the gene diversity also differs. In isolated populations the loci with the least discrimination capacity is TH01, while the one with the highest is FGA. However, in both medium and large populations, the lowest values are in the loci TPOX and the highest in D18S51.

Discussion

As can be seen, forensic parameters, such as PD and H_o , and the analysis of the 13 STR loci are useful tools for individual and paternity testing in a population. However, in this study we compared how these values varied depending on the type of population. It is common to find loss of heterozygosity and presence of similar haplotypes at a higher proportion in small isolated populations because of aboriginal identity preservation.

The aim of this work is to evaluate the forensic efficiency of some statistical parameters. Some forensic cases in isolated populations require modifications in analytical procedures. In daily lab work, it is easy to find single-parent paternity cases. In many cases, a positive result is given in the single-parent situation, but when the other progenitor is included the positive result turns into a negative one. This is even more frequent in case in which the individuals come from an isolated and endogamic population. Table 1 shows a case in which a fatherhood relation was established given a positive result. Later, when the mother of the minor was included in the study as a reference sample, the fatherhood relation turns negative. That is why it is very important to know the ethnic origin of the individuals for creation of genetic databases of isolated populations to better calculate paternity indexes and statistics.

Besides the previously mentioned problems; analyzing the results, it can be deduced that there is a decrease in the value of the power of discrimination (PD) by around 10% in isolated populations compared to medium and large populations. This finding is also confirmed when Y-STRs are studied, as can be seen in the Mayan and Guatemalan Mestizo populations [19], where medium PD values are lower in

isolated populations (Mayan) (0.527) than in big populations (0.654). The H_o value was also lower in isolated population than in medium and large size populations, but in with a less pronounced difference than the previous one.

As can be observed, H_o and PD are lower in all isolated populations than in the other populations. Linguistic and geographical isolation as well as endogamous processes have contributed to the genetic isolation of aboriginal populations. The fact that the isolated populations reside in zones with geographical barriers and rarely interact with other people has limited their genetic pool. This can be a problem in forensic casework.

The use of general genetic frequencies databases derives in erroneous results when samples from isolated populations are studied. There is progressively more genetic information of isolated populations that let the creation of specific genetic frequency tables for each population. If the ethnic origin of the sample is well known and a specific genetic frequency table is available for that population, better statistical results will be given in paternity and criminology testing. The availability of different commercial kit that cover a large number of aSTR markers (IdentifilerPlus, NGM SE, Powerplex 21, PowerplexFusion, Investigator ESSplex SE Plus Kit...) lets a better characterization of isolated population. The more genetic information available of these populations, the better results in the identification of the individuals will be obtained. Therefore, frequency tables can be built with those markers that better characterize the population and give more information (better discrimination power, more heterozygosity, better exclusion power...).

Furthermore, it is important to add to the routine analysis lineage markers; mitochondrial DNA, X-STR markers and Y-chromosome STR.

One of the most important activities toward achieving efficient human identification is that of obtaining reference samples. In contrast with other populations, the study of endogamous populations requires an appropriate family references (direct relatives, parents, or children)

and direct references, such as personal effects or antemortem biological specimens like biopsies and saved bloodstain cards, because of the reduced genetic diversity. In some cases, even when there is only one direct relative, the analyst has to be careful in giving a false positive result. That is why it is important to adapt action protocols when working with these kinds of populations.

Conclusions

It is common knowledge that small and isolated populations have a reduced gene pool which is why each case has to be studied on its own. Action protocols have to be adapted in the daily work with these populations and that is why some recommendations have been proposed in the paper:

- Perform genetic population studies using the highest number of autosomal STR available and determine the ones that give the most genetic information possible and better characterize each population.
- Add complementary lineage information with the use of X and Y STR and mitochondrial DNA.
- Use of appropriate family references (direct relatives, parents, or children) and direct references, such as personal effects or antemortem biological specimens like biopsies and saved bloodstain cards.

Acknowledgements

The authors thank the DNA donors from the city of Granada, Spain, as well as Paraguay, Ecuador, Costa Rica, Brazil, Mexico-Chihuahua, and Guatemala. Furthermore, the authors want to thank the staff of the DNA Laboratory of the National Police of Paraguay, the Criminalistics - DNA Laboratory of the National Police of Ecuador, the OIJ – Forensic Science Laboratory of San Joaquín de Flores (Costa Rica), the Fleury Medical Diagnostic Center - Molecular Biology Section of Sao Paulo (Brazil), the DNA Laboratory of P.G.J. del Estado de Chihuahua (Mexico), and the INACIF (Instituto Nacional de Ciencias Forenses de Guatemala) for kindly providing the DNA samples and for their great effort in collecting information about the origin and ethnicity of each individual in the study. Thanks to all of them, population studies can be performed.

References

1. Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 51: 253-265.
2. Meyer E, Wiegand P, Brinkmann B (1995) Phenotypic differences of STRs in 7 human populations. *Int J Legal Med* 107: 314-322.
3. Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, et al. (2000) Genetic structure of north-west Africa revealed by STR analysis. *European Journal of Human Genetics* 8: 360-366.
4. Zhivotovsky LA, Ahmed S, Wang W, Bittles AH (2001) The forensic DNA implications of genetic differentiation between endogamous communities. *Forensic Sci Int* 119: 269-272.
5. Sharer RJ (1998) La civilización maya. Fondo de Cultura.
6. Coe MD (2005) The Maya.
7. Barrot C, Sánchez C, Ortega M, González-Martín A, Brand-Casadevall C, et al. (2005) Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms. *Int J Legal Med* 119: 111-115.
8. Gazi NN, Eaaswarkhanth M, Ferdous A, Qadri SS, Singh L, et al. (2010) Genetic polymorphisms of 15 autosomal STR loci in three isolated tribal populations of Bangladesh. *Forensic Science International: Genetics* 4: 265-266.
9. Ethnologue: Languages of the World (2009) SIL International, Dallas, Texas.
10. Novokmet N, Pavčec Z (2007) Genetic polymorphisms of 15 AmpFISTR Identifier loci in Romani population from Northwestern Croatia. *Forensic Sci Int* 168:e43-e46.
11. Wu FC, Ho CW, Pu CE, Hu KY, Liu DH (2009) Genetic polymorphisms of 17 Y-chromosomal short tandem repeat loci in Atayal population of Taiwan. *Croat Med J* 50: 313-320.
12. Collins PJ, Hennessy LK, Leibel CS, Roby RK, Reeder DJ, et al. (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: The AmpFtSTR® Identifier® PCR amplification kit. *J Forensic Sci* 49: 1265-1277.
13. Finis CJ (2001) Megaplex STR Analysis from a Single Amplification: Validation of the PowerPlex® 16 System. *Profiles in DNA*: 3.
14. GeneScan Analysis Software Version 3.1 © Copyright 1998, The Perkin-Elmer Corporation.
15. Applied Biosystems (2008) GeneMapper® ID-X Software Mixture Analysis Tool v1.1.
16. Nei M, Roychoudhury AK (1974) Sampling Variances of Heterozygosity and Genetic Distance. *Genetics* 76: 379-390.
17. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.
18. Tereba A (2001) Tools for Analysis of Population Statistics. *Profiles in DNA*.
19. Martínez-González LJ, Saiz M, Álvarez-Cubero MJ, Gómez-Martín A, Álvarez JC, et al. (2012) Distribution of y chromosomal STRs loci in Mayan and Mestizo populations from Guatemala. *Forensic Science International: Genetics* 6: 136-142.