JJBS

Jordan Journal of Biological Sciences

Volume 1, Number 2, Mar. 2008
ISSN 1995-6673
Pages 73 - 76

Virtual Reference Values for STR Genetic Loci Assignment in Forensic Arenas: A Jordanian-Based Study

Raed Khalil [a], Salem Yasin [b,*], Maweih Hamad [b], Ahmad Sharieh [a], Ahmad Al-Jaber [a]

[a] *Computer Science Department, King Abdullah Second for Information Technology, University of Jordan, Jordan*
[b] *Department of Biological Sciences and Biotechnology, Faculty of Science, Hashemite University, Jordan*

Abstract

Short tandem repeat (STR) genetic loci were assigned for use in forensic investigation for the Jordanian population by using a computer program. The program comprised two virtual reference values, the maximum virtual target power of discrimination $(VTP_D)_{max}$ and the maximum virtual observed heterozygosity $(VH_{obs})_{max}$. These two virtual reference values prompted the selection of the TH01, FES/FPS, D18S51, and Penta D STR genetic loci from 17 STR genetic loci used worldwide in forensic applications. The four STR genetic loci showed a combined power of discrimination of 0.999987, indicating that these four STR genetic loci are forensically viable for the Jordanian population. This method of utilizing a computer program to assign genetic loci would reduce the time and cost of DNA typing.

الملخص

لقد تم وضع برنامج محوسب لتحديد أفضل المواقع الجينية الجنائية (STR) لأستخدامها فى المجتمع الأردنى. يضم البرنامج المحوسب قيمتين أفتراضيتين مرجعيتين لتحديد أفضل المواقع الجينية الجنائية. هاتان القيمتين تدعوان القيمة الأفتراضية القصوى للتمييز بين الأفراد $(VTPD)_{max}$ والقيمة الأفتراضية القصوى لتعددية الأليلات التابعة للمواقع الجينية الجنائية $(VHobs)_{max}$. لقد حفزت هذه القيم الأفتراضية أختيار أربعة مواقع جينية جنائية هى: TH01, FES/FPS, D18S51, Penta D من بين سبعة عشر موقعا جينيا جنائيا مستخدمة فى أنحاء العالم. لقد أظهرت المواقع الجينية الأربعة قدرة كبيرة على التمييز بين الأفراد فى المجتمع الأردنى تعادل ٠,٩٩٩٩٨٧. أن أستخدام مثل هذا البرنامج المحوسب يساهم فى خفض تكلفة وتقليل الوقت المستغرق فى تحليل المادة الوراثية DNA.

Keywords: Forensic Science, Genetic Loci, Short Tandem Repeat, Virtual Reference Values;

1. Introduction

The advent of short tandem repeats (STRs) into forensic arenas has revolutionized criminal and legal investigations. The criminal and justice authorities worldwide have utilized the ability of such STRs to solve questionable settings including paternity disputes and human individualization and identification (Budowle et al., 2001; Butler JM, 2006).

Several parameters were established and adopted for the selection of the STR loci as genetic markers in forensic analysis. Such parameters include STR's sequence length (Weber and May, 1989), independent inheritance (Tracey, 2001), polymorphic nature (Weber and May, 1989), heterozygosity which is defined as the probability that two alleles drawn from a population are not identical and amenability to amplification by the polymerase chain reaction (Yamamoto et al., 1997; Lazaruk et al., 1998; Wenz et al., 1998). The polymorphic nature of each STR genetic locus which is a reflection of the STR's allelic window (number of alleles per STR genetic locus) has characterized these short sequences with a highly discriminating power that is useful in human individualization and identification. The discriminating power or power of discriminating (P_D) is defined as the probability that two individuals chosen at random will posses different genotypes for the marker being tested. However, studies have shown demonstrable variation in the allelic windows and also a certain degree of significant differences in the allelic frequency distributions for some of the forensic STR genetic loci among different human populations (Edwards et al., 1992; Gamero et al., 2000; Hamad et al., 2001; Hayes et al., 1995; Lins et al., 1998; Salem et al., 2003; Yasin, 2002). Since the allelic frequencies distribution of the allelic window for each STR genetic locus is implemented in the determination of P_D, it is expected that the variation in the allelic frequencies distribution is also exhibited to a certain degree in their discriminatory power. Thus, discrepancies in the values of P_D of STR genetic loci among human

* Corresponding author. e-mail: omarwyazan@yahoo.com

populations are amenable. Such discrepancies are arguable since they depend on the allelic frequencies distribution in human populations. However, they might still shed some doubt on the efficacy of implementation of the same set of STR genetic loci in the forensic arenas for all human populations. Therefore, the legal implications of such discrepancies should be considered; and prompt to search for a tangible method for determining the efficacy of implementation of a certain genetic locus in a human population.

In this study, the most applicable STRs for forensic investigations in the Jordanian population were selected by developing a computer program. The program utilized two computer-based virtual reference values that depend on the allelic window of each STR genetic locus: the maximum virtual target power of discrimination $(VTP_D)_{max}$ and the maximum virtual observed heterozygosity $(VH_{Obs})_{max}$, i.e., the maximum possible values that could exist in a Hardy-Weinberg population.

2. Materials and Methods

The STR genetic loci raw data implemented in this study were described earlier (Hamad et al., 2001; Salem et al., 2003; Yasin et al., 2002). The data comprised the genotypes and allelic frequency distributions of seventeen STR genetic loci (Table 1).

Figure 1 shows a snapshot of the main menu of the developed software used in this study. The input to program includes various parameters such as population name, population size, sample size, and the experimental STR loci. The developed software generates the virtually observed genotypes of the experimental STR genetic loci within the sample individuals. The results obtained for each STR genetic locus include the allelic frequency distributions, the observed and expected genotypic frequency distributions, and the resultant real and virtual forensic parameters values for all of the experimental STR genetic loci.

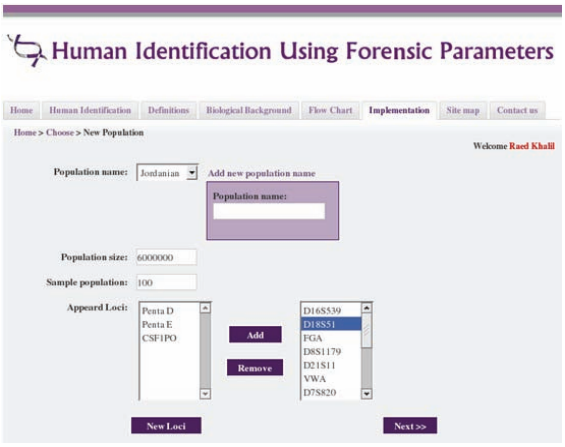

Figure 1: A snapshot of the main menu of the developed software used in this study.

The software comprised two computer-generated virtual reference values, the $(VTP_D)_{max}$ and the $(VH_{Obs})_{max}$. Both reference values require the generation of all virtual genotypes for each genetic locus according to the formula $n(n+1)/2$, where n represents the number of alleles assigned per an STR genetic locus. The $(VTP_D)_{max}$ assumes that in any human population all alleles comprising the allelic window for any STR genetic locus are present in the population. Therefore, the number of all genotypes that could possibly exist or emerge in the population can be determined and listed.

The computer software would randomly generate observed numbers for each genotype; and then, determine the allelic frequencies distribution for each allele in the STR allelic window.

Table 1: Virtual Reference Values for the Seventeen STR Genetic Loci.

STR Locus	*Power of Discrimination (P_D)*			*Observed Heterozygosity (H_{Obs})*		
	Actual P_D	$(VTP_D)_{max}$	*p* value	Actual H_{Obs}	$(VH_{Obs})_{max}$	*p* value
TH01	0.94360	0.98437	0.151	0.75	0.833	0.0751
TPOX	0.8284	0.98138	0.0003*	0.45	0.81	0.0000*
vWA	0.90800	0.98845	0.0139*	0.84	0.856	0.3765
Penta D	0.95700	0.99360	0.097	0.87	0.886	0.3650
Penta E	0.97640	0.99583	0.2417	0.73	0.928	0.0001*
CSF1PO	0.85320	0.98670	0.0006*	0.85320	0.98670	0.0006*
FGA	0.95300	0.99833	0.039*	0.84	0.949	0.0065*
FES/FPS	0.87640	0.95213	0.0572	0.74	0.713	0.3344
F13A1	0.91660	0.98848	0.0178*	0.78	0.864	0.0610
D3S1358	0.89240	0.98664	0.0057*	0.84	0.851	0.4149
D5S818	0.89860	0.98135	0.0146*	0.72	0.815	0.0567
D7S820	0.92120	0.98437	0.0365*	0.80	0.852	0.1666
D8S1179	0.93480	0.99416	0.0244*	0.76	0.898	0.0051*
D13S317	0.89200	0.98131	0.0102*	0.79	0.822	0.2839
D16S539	0.88260	0.98137	0.0061*	0.82	0.824	0.4706
D18S51	0.96180	0.99727	0.078	0.88	0.934	0.0951
D21S11	0.95040	0.99726	0.0392*	0.80	0.938	0.0021*

*: Significant level of difference

The sample size was virtually assigned at 1000 virtual individuals. These allelic frequencies are used to calculate the P_D at each random generation of observed number for each genotype of an STR genetic locus. The computer software would then determine the $(VTP_D)_{max}$ reference value that could result. This $(VTP_D)_{max}$ reference value is used as a reference target value to determine the applicability of an STR genetic locus in forensic arenas in the Jordanian population. The raw data of the power of discrimination based upon the allelic frequency distributions present in the Jordanian population for each STR genetic locus were compared with the $(VTP_D)_{max}$ reference value. Based on all virtual genotypes, the software determined the $(VH_{Obs})_{max}$ reference values for all of the seventeen STR genetic loci. The $(VTP_D)_{max}$ and the $(VH_{Obs})_{max}$ reference values for the seventeen STR genetic loci were considered to be the target values for determining the applicability of any genetic STR locus in human identification settings in Jordan.

Statistical analysis between the raw data and the virtual data for all the seventeen STR genetic loci was carried out using the STATISTICA software for Windows, 1995 version (StatSoft, Tulsa, OK, USA), where the test for difference between two percentages computes the significance level for the difference between two percentages; both one-sided and two sided tests. The *p*-level was computed based on the t-value for the respective comparison according to the following formulas:

$$| t | = \sqrt{[(N1+N2)/(N1+N2)]x[P1-P2]/ \sqrt{(Pxq)}}$$

Where,

$$P=(P1xN1+P2xN2)/(N1+N2)$$
$$q = 1- P$$

The degrees of freedom are computed as:

$$N1 + N2 - 2$$

3. Results

In view of the computer-generated $(VTP_D)_{max}$ values (Table 1), the computer software was able to randomly generate the virtual genotypes and allelic frequencies for all of the STR genetic loci. The allelic frequencies were used to generate the $(VTP_D)_{max}$ values for each genetic STR locus; while the $(VH_{Obs})_{max}$ values were calculated using the computer-generated virtual genotypes.

The $(VTP_D)_{max}$ values varied from 0.99833 for the FGA genetic locus to 0.95213 for the FES/FPS genetic locus. These indicated that all of the seventeen genetic STR loci possess high virtual target discriminatory power that can be applied in determining the efficacy of each respective genetic locus for forensic applications. Furthermore, the results showed that the $(VH_{Obs})_{max}$ values for the seventeen STR genetic loci varied from 0.949 for the FGA STR genetic locus to 0.713 for the FES/FPS STR genetic locus.

In order to evaluate the efficacy of forensic implementation of each of the seventeen STR genetic loci in the Jordanian population, the previously determined allelic and genotypic windows for the Jordanian population were used to recalculate the various forensic parameters. They include the actual P_D and H_{Obs} values for each STR genetic locus. These parameters were compared with previously published data, particularly, with their respective $(VTP_D)_{max}$ and $(VH_{Obs})_{max}$ values for each STR genetic locus (Table 1).

Data analysis showed that all previously published forensic parameters are similar to these generated by the computer software (Hamad et al., 2001; Yasin, 2002; Salem et al., 2003). In this work, attention was paid for the actual P_D and H_{Obs} values for all of the seventeen STR genetic loci. The actual P_D values varied from 0.97640 for the Penta E genetic locus to 0.82840 for the TPOX genetic locus. The actual H_{Obs} values varied from 0.88 for the D18S51 STR genetic locus to 0.45 for the TPOX STR genetic locus.

Statistical comparison of the $(VTP_D)_{max}$ values with the actual P_D values showed significant differences ($p < 0.05$) at twelve genetic STR loci, namely, TPOX, vWA, CSF1PO, FGA, F13A1, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, and D21S11 (Table 1). Absence of significant differences was demonstrated for the TH01, Penta D, Penta E, FES/FPS, and D18S51 STR genetic loci ($p > 0.05$). Furthermore, comparison between the $(VH_{Obs})_{max}$ values and the actual H_{Obs} values (Table 1) showed statistically significant differences ($p < 0.05$) at six STR genetic loci, namely, TPOX, Penta E, CSF1PO, FGA, D8S1179, and D21S11. No significant differences ($p > 0.05$) were shown for the TH01, vWA, Penta D, FES/FPS, F13A1, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, and D18S51 STR genetic loci data.

4. Discussion

Table 1 indicates that the actual P_D values for the seventeen genetic loci are high enough to enable them to show a high discriminatory power among Jordanian individuals. Nevertheless, twelve of these STR genetic loci could be recommended for exclusion from the STR genetic loci panel used for forensic investigations in the Jordanian population because there was a significant difference between the virtual and actual P_D values for these loci. Such exclusion reduces the number of STR genetic loci being used to a panel of five loci that includes the TH01, Penta D, Penta E, FES/FPS, and D18S51. The combined power of discrimination for these five genetic loci was calculated at 0.999999692 (Combined P_I = 0.000000308). Thus, combining these five STR genetic loci into a single panel should yield satisfactory levels of population resolution and individual identification in Jordanians considering the small size of the Jordanian population that approximates 5.65 million according to the latest Jordanian census bureau data (http://www.dos.gov.jo).

The presence of significant level of differences between the $(VH_{Obs})_{max}$ values and the actual H_{Obs} values of the TPOX, Penta E, CSF1PO, FGA, D8S1179, and D21S11 STR genetic loci indicated that the exhibited level of observed heterozygosity for these six STR genetic loci in the Jordanian population is theoretically not acknowledgeable. Accordingly, these STR genetic loci can be excluded from any forensic STR set used for human

identification or individualization in the Jordanian population. Furthermore, the data demonstrated that the TH01, vWA, Penta D, FES/FPS, F13A1, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, and D18S51 STR genetic loci exhibit high heterozygosity levels, hence high genetic variability in the Jordanian population. This indicates their applicability in forensic investigation in Jordan. These eleven STR genetic loci show an exceptionally high combined P_D of 0.99999999999891 (Combined $P_I = 1.0914x10^{-12}$).

Combining both of the virtual reference values of $(VTP_D)_{max}$ and $(VH_{Obs})_{max}$ can greatly enhance the selection of the forensic STR genetic loci panel used for human identification and individualization in a population. In Jordanian population, the forensic STR genetic loci panel could comprise four STR genetic loci, namely, the TH01, Penta D, FES/FPS, and D18S51. The combined Power of Discrimination for these four loci was calculated at 0.999987 (Combined $P_I = 0.000013$). This is a relatively high discriminatory power considering the population size of Jordan.

In this work, we established a virtual method that is able to determine the fitness of and assign STR genetic loci into a low-cost, time-saving, and a highly discriminatory forensic STR panel. However, it remains for the national forensic authorities in the world to evaluate the applicability of any forensic genetic loci in their respective populations taking in consideration the population size and possible exchanges with neighbouring countries. It is possible that a higher population size or mosaic population could require increasing the number of STR genetic loci to achieve significantly higher discriminatory power.

References

[1] Budowle B, Shea B, Niezgoda S, Chakraborty R (2001) CODIS STR loci data from 41 sample populations. J Forensic Sci 46(3):453-489.

[2] Butler JM (2006) Genetics and genomics of core STR loci used in human identity testing. J Forensic Sci 51(2):253-265.

[3] Edwards A, Hammond HA, Jin L, Caskey CT and Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241-253.

[4] Gamero JJ, Romero JL, Gonzalez JL, Arufe MI, Cuesta MI, Corte-Real F, Carvalho M, Anjos MJ, Vieira DN and Vide MC (2000) A study on ten short tandem repeat systems: African immigrant and Spanish population data. Forensic Sci Int 110:167-177.

[5] Hamad M, Yasin SR, Elkarmi A (2001) Polymorphism of HUMvWA31, HUMTH01, HUMF13A1 and HUMFES/FPS STR genetic loci in Jordanians. Korean J Genetics 23(2):157-161.

[6] Hayes JM, Budowle B and Freund M (1995) Arab population data on the PCR-based loci: HLA-DQA1, LDLR, GYPA, HBGG, D7S8, Gc, and D1S80. J Forensic Sci 40(5):888-892.

[7] Lazaruk K, Walsh PS, Oaks F, Gilbert D, Rosenblum BB, Menchen S, Scheibler D, Wenz HM, Holt C and Wallin J (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. Electrophoresis 19:86-93.

[8] Lins AM, Micka KA, Sprecher CJ, Taylor JA, Bacher JW, Rabbach DR, Bever RA, Creacy SD and Schumm JW (1998) Development and population study of an eight-locus short tandem repeat (STR) multiplex system. J Forensic Sci 43(6):1168-1180.

[9] Salem K, Yasin S, Elkarmi A, Hamad M, Jaran A (2003) Jordanian population data on five STR forensic loci: D16S539, TPOX, CSF1PO, Penta D, and Penta E. Legal Med 5:251-252.

[10] Tracey M (2001) Short tandem repeat-based identification of individuals and parents. Croat Med J 42:233-238.

[11] Weber JL and May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Legal Med 44:388-396.

[12] Wenz H, Robertson JM, Menchen S, Oaks F, Demorest DM, Scheibler D, Rosenblum BB, Wike C, Gilbert DA and Efcavitch JW (1998) High-precision genotyping by denaturing capillary electrophoresis. Genomes Res 8:69-80.

[13] Yamamoto T, Kojima T, Nozawa H, Huang X-L, Ohtaki H, Uchihi R, Tamaki K and Katsumata Y (1997) A rapid typing system at three STR loci from bloodstains using a simple DNA extraction kit and capillary electrophoresis. Jpn J Legal Med 51:396-400.

[14] Yasin SR (2002) Allele frequencies at nine PCR-based STR loci in Jordanians. Korean J Genetics 24(4):327-334.