Research Article   **Open Access**

Effect of DNA Profile Size, Reference Population Database, and Parents Availability on Parentage Testing in Consanguineous and Endogamous Populations: The Lebanese Case

Ansar El Andari[1,2#], Ayman Daouk[1#] and Issam Mansour[1*]

[1]Department of Laboratory Science and Technology, Faculty of Health Science, American University of Science and Technology, Beirut, Lebanon

[2]School of Criminal Justice, Faculty of Law and Criminal Justice, University of Lausanne, Lausanne, Switzerland

[#]These authors contributed equally to this work.

[*]Corresponding author: Issam Mansour, DNA Testing and Research Advisor, Department of Laboratory Science and Technology, Faculty of Health Science, American University of Science and Technology, Beirut, Lebanon, Tel: +961218716230; E-mail: fhs@aust.edu

Abstract

Short tandem repeats (STR) analysis is currently the gold-standard method for DNA profiling in parentage testing where the probabilistic analysis is affected by the DNA profile size, the STR frequency dataset, and the availability of all the individuals involved in the case. Lebanon is a relative small country with high rates of endogamous and consanguineous marriages; factors that could challenge the power of discrimination of the DNA profiles and introduce uncertainty to parentage testing interpretations. Since there is no legislation concerning kinship testing in Lebanon, relationship testing laboratories follow standards of their own choice, the majority of parentage testing is performed using 15 STR systems only, and some laboratories still use Caucasian STR allele frequencies as reference population datasets. In addition, there is a remarkable increase in Duo parentage testing requests. The present study aimed at assessing the effect of the number of tested STR loci 15 versus 23, and the effect of using Lebanese versus Caucasian STR allele frequencies as reference population dataset in 422 Trio and 271 Duo parentage cases. Results showed that performing paternity testing using DNA profiles of 15 STR systems in the Lebanese population increased the risk of misinterpretations, whereas increasing the DNA profile size to 23 STR markers using the population representative STR allele frequency helped remarkably in lowering uncertainties, and probable misinterpretations, especially in Duo parentage cases.

Keywords: DNA profiling; Kinship testing; Parentage testing; Paternity; Maternity; Consanguinity; Endogamy; Lebanon

Introduction

Lebanon is a relatively small country that has been shaped by many civilizations with profound effects at the cultural, religious and genetic levels [1]. As a result, the Lebanese population is at present characterized by more than 18 religious communities that are not evenly distributed geographically, live in a relative seclusion, and have different religious and cultural practices. Consequently, marriages take place within the same religious community at an average rate of 88% [2], and between first degree relatives with rates ranging from 31% to 36% [3,4]; factors which helped shaping the genetic structure of the Lebanese communities [5].

Given that Lebanon has relatively high endogamous and consanguineous marriage rates, the likelihood of observing matching allele types among different individuals is high [6, 7], in addition to a loss in systems heterozygosity. In parentage testing, when the involved individuals belong to the same subpopulation, falsely alleged parents are less likely to exhibit system allele mismatches with the child's genotype, than when the population is not subdivided. Consequently, paternity studies in the Lebanese population could be prone to significant uncertainties.

Besides, estimating STR allele frequencies in a given population is an essential prerequisite for any DNA interpretation. However, several studies showed that the population source of STR allele frequencies do not have significant impact in parentage testing calculations, where the use of different population frequency databases could lead to the same result [8-11]. Yet a subdivided population does not follow a relative panmixia [12] as its individuals would not randomly mate; therefore, the STR allele frequencies could differ among the different subpopulations or between a subpopulation and the total representative population allele frequencies. The Lebanese population shows stratification with inbreeding practices that could affect STR allele frequencies, not only in comparison to the Caucasian population, but also among different Lebanese religious sub communities.

STR allele frequencies that are in Hardy-Weinberg equilibrium with low coefficient of coancestry (F_{st}), such as the Caucasian STR allele frequencies are still used as reference in certain Lebanese DNA testing laboratories, although Lebanese STR allele frequencies are available; this could probably be a source of statistical bias [13,14]. Also, the Lebanese social characteristics challenge the discrimination power of DNA profiles as to the number of included STR systems, which are often 15. Indeed, many studies have discussed the effect of underestimating the inbreeding coefficient in DNA interpretations and questioned the number of used genetic markers, where these reports showed that these conditions may introduce uncertainties in the DNA interpretations [12, 15-22]. Cases where 21 STR genetic markers are deficient to prove paternity, as well as false paternity positivities in Duo cases, have been documented in populations experiencing random mating [23-27].

Lebanon lacks legislation concerning kinship testing; therefore, relationship testing laboratories follow recommendations and standards of their choice, with recommendation of positive parentage probabilities (W) ranging from 99.9% to 99.999% [28].

The present study aims at evaluating the effect of using population specific STR allele frequencies and assessing the effect of the DNA profile size and both parents availability on parentage investigations in populations with recurrent inbreeding practices as is the Lebanese population.

Methodology

Samples collected were either buccal swabs or peripheral blood using EDTA tubes. Genomic DNA was extracted from whole blood using the salting out method, whereas organic DNA extraction, by Phenol Chloroform Isoamylalcohol (PCI) protocol, was used for buccal swabs. The DNA extracted from each sample was quantified using Nanodrop.

Approximately 1 ng of the target DNA was amplified using two commercial typing kits, PowerPlex1 16 HS and PowerPlex1 ESI 17 fast (Promega Corporation, Madison, USA), according to manufacturer's instructions. These two kits combined amplify a total of 23 genetic markers (D3S1358, D22S1045, THO1, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, VWA, D8S1179, TPOX, FGA, D19S433, D2S1338, D1S1656, D10S1248, D2S441, D12S391 and SE33). The PX2 thermohybaid (Thermo) was used for PowerPlex ESI 17 fast amplification, whereas the AB GeneAmp PCR System 9700 (Applied Biosystems) was used for PowerPlex 16 HS amplification. Inconclusive cases were tested with the PowerPlex® CS7 for an additional five systems.

Genotypic data was obtained using ABI 3130 Genetic Analyzer using Performance Optimized Polymer (POP) 7 and 4 cm-36 cm capillary arrays (Applied Biosystems, California, USA). Internal Lane Standards (ILS) 600 and 500 were used for sizing the determinations of amplified fragments using PowerPlex 16 HS and PowerPlex ESI 17 fast, respectively. Data collection was achieved using the Data Collection Software v 3.0 and fragment analysis was performed using GeneScan Analysis Software v 4.1 from Applied Bio systems. Size calling was obtained using allelic ladders provided in the kits. The obtained data was then imported to an in-house Forensic Information Management System (FIMS) to calculate the frequencies.

422 Trio parentage cases were reanalysed, consisting of 211 paternity and 211 maternity Trio investigations. Duo parentage cases were simulated out of these Trio cases totalling 422 Duo simulations. In addition, 271 Duo parentage family cases were evaluated, consisting of 244 paternity and 27 maternity Duo investigations; making a total of 693 Duo investigations analysed.

Each paternity case was analysed four times, (i) Lebanese STR frequencies-15 STRs (ii) Caucasian STR allele frequencies-15 STRs (iii) Lebanese STR frequencies-23 STRs (iv) Caucasian STR allele frequencies-23 STR systems. The percentage of paternity (POP) calculated using Essen-Möller formula (based on a prior probability of 0.5) was compared to assess the difference between each combinational analysis of the same case.

Comparisons were performed between the results obtained for: (i) Duo 15 STRs versus Duo 23 STRs (ii) Trio 15 STRs versus Trio 23 STRs (iii) Trio 15 STRs versus Duo 23 STRs (iv) Trio 15 STRs versus Duo 15 STRs (v) Trio 23 STRs versus Duo 23 STRs, all using Lebanese

STR allele frequencies; and (vi) Duo 23 STRs with the Lebanese STR allele frequencies versus the Caucasian STR allele frequencies (vii) Trio 15 STRs with the Lebanese STR allele frequencies versus the Caucasian STR allele frequencies.

Results

DNA profiles using 15 versus 23 STR markers

Duo cases using 15 STR versus 23 STR markers: When Duo cases were tested using 15 and 23 STR markers, the calculated parentage probability (POP or W) based on the Essen-Möller formula changed showing clear positive parentage probabilities in most of the cases with the increase in the DNA profile size from 15 to 23 STRs (Table 1).

	15 STR Systems	23 STR Systems
POP ≥ 99.999%	447	684
99.99% ≤ POP<99.999%	164	2
99.9% ≤ POP<99.99%	69	6
99% ≤ POP<99.9%	11	1
POP<99%	2	0

Table 1: Parentage probability (W) using 15 *vs* 23 STR systems in 693 Duo families (n= 693).

For all cases, the POP significantly increased when the number of tested STR systems increased (data not shown). With 15 STRs, 164 cases out of 693 showed a POP between 99.99% and 99.999%. These cases drastically decreased to 2 out of 693 when 23 STRs when used. Also, with a DNA profile of 15 STR systems, 69 cases out of 693 showed a POP between 99.9 % and 99.99%, out of which only 6 cases showed a between 99.9% and 99.99% when tested using 23 STR. Additionally, 10 out of the 11 cases that were tested using 15 STR systems showing a POP between 99% and <99.9%, showed a POP>99.9% when using 23 STR systems. Out of those 10 cases, 8 showed a POP>99.99% and 2 cases showed a POP>99.999%. Only 1 case out of 693 investigations (0.14% of the cases) remained with a POP<99.9% with 23 STR systems.

A total of nine cases remained with a POP lower than 99.999% when tested with 23 STR systems; two solutions were attempted: i) increasing the number of tested systems to 28 STR systems, ii) including the missing parent's profile when available. Results are shown in the following Table 2.

Case	Duo	Duo	Trio
	23 STRs	28 STRs	15 STRs
DF05035	97.4	99.9	N/A
DF12003b	95.7	99.7	N/A
DF11068c	99.995	99.999	N/A
DF11093n	99.998	99.99999	99.999
DF07053a	99.8	99.8	100
DF07053b	99.7	99.9	100

DF10063a	99.97	99.99	100
DF10046a	99.998	99.999	99.9999
DF10046b	99.7	99.9998	100

Table 2: Percentage of paternity (POP) for cases with inconclusive parentage results when treated under different scenarios

For all the nine cases, the number of tested STR was increased by adding 5 more systems (using CS7 STR kit from Promega) to have a total of 28 STR loci. 4 cases out of the 9 showed a POP ≥ 99.999%, whereas 5 cases remained below this threshold. However, when the second parent was available for the analysis, all cases showed a POP higher than 99.999 using 15 STR systems (in its Trio scenario).

Trio cases using 15 STR versus 23 STR markers: Results using 15 STR systems for the analysis of the parentage cases showed that only 12 cases were below 99.999%. All these cases showed a POP higher than 99.999% when the number of STRs was increased to 23 markers (Table 3).

	Trio 15 STRs *vs* Duo 15 STRs		Trio 15 STRs *vs* Duo 23 STRs		Trio 15 STRs *vs* Trio 23 STRs		Trio 23 STRs *vs* Duo 23 STRs	
n=422	Trio	Duo	Trio-15	Duo-23	Trio-15	Trio-23	Trio	Duo
POP ≥ 99.999%	410	298	410	412	410	422	422	412
99.99% ≤ POP<99.999%	6	86	6	4	6	0	0	4
99.9 % ≤ POP<99.99%	4	30	4	3	4	0	0	3
99% ≤ POP<99.9%	1	7	1	3	1	0	0	3
POP<99%	1	1	1	0	1	0	0	0

Table 3: Percentage of parentage or parentage probability (W) when using different scenarios as of DNA profile size and parent availability in Trio Lebanese cases.

For all cases, the POP significantly increased when the number of tested STR systems increased (data not shown). In addition with 15 STRs, 11 cases out of 422 showed a POP between 99.9% and 99.999%, while 1 case showed a POP below 99%. All these cases showed a POP ≥ 99.999% when tested using 23 STRs.

Trio versus Duo scenarios

Trio cases versus their simulated duo scenario using 15 STR markers: In Trio investigations using DNA profiles of 15 STRs, only 12 cases out of 422 cases showed POP values below 99.999% (<3% of the cases), while in Duo cases, the number was 124 cases out of 422 cases (~29% of the cases). In Duo cases, the percentage of parentage decreases due to the absence of one of the parents' profile. There is a remarkable increase in the number of cases below 99.999% in Duo scenarios compared to that of the Trio cases (Table 3) when tested using 15 STR systems.

Trio cases using 15 STRs versus their simulated duo scenario using 23 STR markers: The number of cases below 99.999% in the Duo cases using 23 STR systems is less than that of Trio cases using 15 STR systems (Table 3). 10 Duo cases tested using 23 STR systems showed a POP below 99.999% while 12 Trio cases tested using 15 STR systems, showed a POP below 99.999%, 2 of them being below 99.9% which is considered as inconclusive according to the AABB inclusion threshold [26]. This result showed the importance of increasing the number of STR loci to 23 markers, especially when dealing with Duo (fatherless or motherless) cases where a compensatory factor is need for the absence of one of the parents' profiles.

Trio cases versus their simulated duo scenario using 23 STR markers: Even with 23 systems, 10 of the Duo cases showed a POP lower than 99.999% while all Trio cases showed a POP higher than 99.999% with the same number of STR systems. This indicated that all of the cases were solved when both parents were present and maximum number of STRs was used.

Lebanese versus Caucasian STR allele frequencies

Parentage investigations with 23 STRs markers using Lebanese STR allele frequencies versus the caucasian STR allele frequencies: The number of cases having a POP higher than 99.999% using Lebanese STR allele frequencies was 1086 out of 1115 cases, while it was 1095 when using Caucasian STR allele frequencies. Also, while comparing cases having POP <99.999 between the Lebanese and the Caucasian frequencies, 29 cases showed a POP below this threshold with the Lebanese frequencies, while 20 cases showed a POP below 99.999 when using the Caucasian frequencies. Since the Lebanese STR allele frequencies are representative of the Lebanese population, this means that results using the Lebanese frequencies are more representative than those using the Caucasian frequencies (Table 4).

	Lebanese Fr.	Caucasian Fr.
POP ≥ 99.999%	1086 / 1115	1095 / 1115
99.99% ≤ POP<99.999%	12 / 1115	12/ 1115
99.9% ≤ POP<99.99%	10/ 1115	8/ 1115
99% ≤ POP<99.9%	4 / 1115	0 / 1115
POP<99%	3 / 1115	0 / 1115

Table 4: Percentage of parentage probability (W) with 23 STR systems for Lebanese parentage investigations using Lebanese *vs.* Caucasian STR allele frequencies.

Discussions

STR analysis has shown to be one of the most reliable molecular biology tools in forensic caseworks and parentage testing owing to its ease of use and a high degree of polymorphism among individuals. The power of discrimination of an allele within each STR marker depends on its frequency in a given population, which implies the need for each population to estimate its own STR allele frequencies dataset.

Determining the number of tested markers to establish a DNA profile also affects the statistical power of analysis; increasing the number of tested STR loci has a positive effect on the degree of certainty. In parentage testing, when dealing with motherless or fatherless (Duo) cases, the uncertainty of statistical analysis increases relatively to the cases where the paternal, maternal, and child profiles are available (Trio cases).

There is no local standard legal inclusion threshold, therefore, cases are interpreted based on different international set recommendations like the American Association of Blood Bank recommended inclusion threshold ($\geq$ 99.9%) or be more stringent in the analysis and adopt the French and German recommended inclusion threshold ($\geq$ 99.999%).

The present study evaluates the effect of three main factors: (a) the effect of the number of STR markers, (b) the effect of the availability of both parents in the study (Duo versus Trio), and (c) the effect of choosing population specific STR allele frequencies

When comparing cases tested with 15 STR systems versus 23 STR systems, we obtained a remarkable higher percentage of parentage when using profiles composed of 23 STR systems than with 15 STRs, whether the cases were Duo or Trio, which emphasizes on the need of testing 23 STR systems when dealing with parentage cases, especially when the cases are Duo (motherless or fatherless cases). Duo cases showing a POP below 99.999% were solved by increasing the number of tested STR systems from 15 to 23. This fact assures the need of increasing the number of tested STR systems especially with challenging Duo cases.

When comparing Trio cases with their Duo simulations and using the same number of tested STR systems (Table 3), the results of Trio cases were much more certain than those of Duo simulation cases. When comparing the Trio cases with 15 STR systems versus their Duo simulation with 23 STR systems, results showed that the absence of one of the parents profile was successfully compensated with the increase of the number of tested STR loci (Table 3). This is due to the fact that the presence of all the parentage testing components (father, mother and child) have a remarkable effect on the level of certainty for any parentage testing regardless of the number of tested STR systems.

The difference between the Lebanese STR allele frequencies and the Caucasian STR allele frequencies was clearly shown during the assessment of the Lebanese STR allele frequencies [13].

Comparisons showed that some of the limited cases became inclusion when the Caucasian frequencies were used. These few cases were below the threshold (AABB) when the Lebanese frequencies were used and became above this threshold when the Caucasian frequencies were used. This means that the allele shared between the child and one of the parents can be less frequent in the Lebanese than the Caucasian population. Since Lebanese STR frequencies are representative to the Lebanese population, thus it is highly recommended to all parentage testing laboratories in Lebanon the use of the Lebanese STR systems to minimize biased results when the Caucasian STR allele frequencies are used, especially when accepting Duo parentage cases.

Conclusions

The present study showed that increasing the number of tested STR systems, from 15 to 23, gave more accurate calculations and results when dealing with Duo cases. Therefore, it is strongly recommended for all parentage profiling laboratories to test 23 STR systems with all Duo cases (motherless or fatherless) in order to have a degree of certainty and confidence similar to that when testing 15 STR systems for Trio cases. When using 15 STRs in Trio cases, only 2% of cases showed inconclusive results. So when testing Trio cases, we have the choice of using either 15 or 23 STR systems.

Also, it is highly recommended for all DNA profiling laboratories in Lebanon to rely on the Lebanese STR allele frequencies to minimize the bias and maximize the degree of certainty when dealing with parentage testing since the Caucasian STR allele frequencies are not representative for the Lebanese population.

Acknowledgements

References

1. Salibi KS (1965) Modern History of Lebanon. (1stedn), Weidenfeld & Nicolson, London

2. Klat M, Khudr A (1986) Religious endogamy and consanguinity in marriage patterns in Beirut, Lebanon. Biodemography and Social Biology 33:138-145.

3. Kanaan ZM, Mahfouz R, Tamim H (2008) The Prevalence of Consanguineous Marriages in an Underserved Area in Lebanon and Its Association with Congenital Anomalies. Genet Testing 12: 367-372.

4. Barbour B, Salemeh P (2009) Consanguinity in Lebanon: Prevalence, distribution and determinants. J Biosoc Sci 41: 505-517.

5. Mansour I, Delague V, Cazeneuve C, Dode C, Chouery E, et al. (2001) Familial Mediterranean fever in Lebanon: mutationspectrum, evidence for cases in Maronites, Greek orthodoxes, Greek catholics, Syriacs and Chiites and for an association between amyloidosis and M694V and M694I mutations. Eur J Hum Genet 9: 51-55.

6. Ayres K (2000) Relatedness testing in subdivides populations. Forensic Sci Int 114: 107-115.

7. Abbas S, Mourad L, Mansour I (2018) Evaluation of Sibling-Ship Analysis in Secluded Lebanese Villages with Increased Endogamous Mating Patterns. J Forensic Investigation 6: 4.

8. Fernandes AT, Goncalves R, Brehm A (2004) Databases: The real importance in paternity testing. Int Congr Ser 1261: 463-464.

9. Moroni R, Gasbarra D, Arjas E, Lukka M, Ulmanen I (2011) Effects of reference population and number of STR markers on paternity testing. J Forensic Res 2.2.

10. Thomson JA, Pillotti V, Stevens P, Ayres K, Debenham PG (1996) Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. Forensic Sci Int 100: 1-16.

11. Lee LD, Fairley M, Lambert JA, Evett IW (1996) Validation of a frequency database for four STR loci for use in casework in the Strathclyde police forensic science laboratory. Forensic Sci Int 79: 43-48.

12. Balding D, Bishop M, Cannings C (2007) Handbook of Statistical Genetics. (1stedn), John Wiley & Sons, UK.

13. El Andari A, Othman H, Taroni F, Mansour I (2013) Population Genetic Data for 23 STR Markers from Lebanon. Forensic Sci Int Genet 7: 108–113.

14. ChoueryE, Coble MD, Strouss KM, Saunier JL, Jalkh N, et al. (2010) Population genetic data for 17 STR markers from Lebanon. Leg Med 12: 324-326.

15. Goodwin W, Ballard D, Simpson K, Thacker C, Syndercombe Court D, et al. (2003) Case study: Paternity testing-when 21 loci are not enough. Int Congr Ser 1261: 460-462.

16. Jacewicz R, Berent J, Prosniak A, Kadlubek M, Szram S (2004) The evaluation of the IDENTIFILER system in paternity testing in Poland. Int Congr Ser 1261: 538-540.

17. Babol-Pokora K, Jacewicz R, Papinski, Szram S (2006) Identifiler system as an inadequate tool for judging motherless paternity cases. Int Congr Ser 114: 462-464.

18. Babol-Pokora K, Jacewicz R, Papinski, Szram S (2006) Is SGM Plus the sufficient system for paternity testing?. Int Congr Ser 1288: 450-452.

19. Moroni R, Gasbarra D, Arjas E, Lukka M, Ulmanen I (2008) Effects of reference population and number of STR markers on paternity testing. Forensic Sci Int Genet 1: 654-655.

20. Evett IW, Weir BS (2008) Intepreting DNA evidence. (1stedn), Sinaue associates, INC, pp: 95.

21. El Andari A and Mansour I (2013). Is he or is he not the father? Pros, Cons effect of following standardized interpretations. Forensic Sci Int Genet 4:e186-e187.

22. El Andari A and Mansour I (2017) Effect of motherless paternity cases on the interpretation of parentage investigations in a population with recurrent inbreeding practices. J Forensic Res 8: 395.

23. Sanchez D, Gonzalez-Andrade F, Bolea M, Martinez Jarreta B (2008) False inclusion in a deficient paternity case with two alleged fathers. Forensic Sci Int Genet 1: 525-527.

24. Von Wurmb-Schwark N, Malyusz V, Simeoni E, Lignitz E, Poetsch M (2006) Possible pitfalls in motherless paternity analysis with related putative fathers. Forensic Sci Int 159: 92-97.

25. Poetsch M, Ludcke C, Repenning A, Fischer L, Malyusz V, et al. (2006) The problem of single parent/child paternity analysis-Practical results involving 336 and 348 unrelated men. Forensic Sci Int 159: 98-103.

26. Jacewicz R, Berent J, Prosniak A, Dobosz T, Kowalczyk E, et al. (2004) Non-exclusion paternity case with a triple genetic incompatibility. Int Congr Ser 1261: 511-513.

27. Hou J, Tang H, Liu Y, Hou Y (2008) How many markers are enough for motherless cases of parentage testing. Forensic Sci Int Genet 1: 649-651.

28. American Association of Blood Bank (AABB) (2009) Standards for Relationship Testing Laboratories (9th edition) USA.