JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, July 2012, Vol. 57, No. 4
doi: 10.1111/j.1556-4029.2012.02095.x
Available online at: onlinelibrary.wiley.com

CASE REPORT

CRIMINALISTICS

Mark Barash,[1,] M.Sc.; Ayeleth Reshef,[1] M.Sc.; Lev Voskoboinik,[1] M.Sc.; Ashira Zamir,[2] M.Sc.; Uzi Motro,[3] Ph.D.; and Ron Gafny,[1] Ph.D.*

A Search for Obligatory Paternal Alleles in a DNA Database to Find an Alleged Rapist in a Fatherless Paternity Case

ABSTRACT: A sexual assault case resulted in a pregnancy, which was subsequently aborted. The alleged father of the fetus was unknown. Maternal and fetal types were obtained using the 11-locus AmpFℓSTR® SGM Plus® kit. The national DNA database was searched for the paternal obligatory alleles and detected two suspects who could not be excluded as father of the male fetus. Additional typing using the AmpFℓSTR® Minifiler™ kit, containing three additional autosomal loci, was not sufficient to exclude either suspect. Subsequent typing using the PowerPlex® 16, containing four additional loci, and Y-Filer™ kits resulted in excluding one suspect. Searching a database for paternal obligatory alleles can be fruitful, but is fraught with possible false positive results so that finding a match must be taken as only preliminary evidence.

KEYWORDS: forensic science, forensic DNA analysis, sexual assault, paternity, DNA database, AmpFℓSTR® SGM Plus®, AmpFℓSTR® Minifiler™, PowerPlex® 16, AmpFℓSTR® YFiler®

A young woman arrived at the hospital and requested an abortion procedure for a pregnancy that she claimed was a result of a sexual assault. The victim stated that the sexual offense had taken place about 12 weeks previously and that during the attack she did not manage to see the offender. Following her testimony, an abortion procedure was carried out. The abortion material and the victim's reference sample were sent for analysis in the forensic biology laboratory, and the profiles of the victim and the fetus were obtained using the AmpFℓSTR® SGM Plus® kit (Applied Biosystems, Foster City, CA).

Because the offender was unknown and because at the assumed attack area, three more unresolved sexual assaults were reported, it was decided to search the Israeli national DNA database for probable suspects who could not be excluded as alleged fathers of the aborted fetus. The obligatory paternal alleles of the alleged father at seven loci were submitted for search in the DNA database. At the other three loci, the fetus and the mother were both identical heterozygotes, so a single obligatory parental allele could not be established.

The search of 55,000 profiles in the reference DNA database resulted in three potential suspects. Subsequent comparison to the mother's and aborted fetus profiles at all of AmpFℓSTR® SGM Plus® loci excluded one of the three suspects. One of the remaining suspects lived in the assumed area of attack ("suspect 1"), while the second suspect ("suspect 2") lived approximately 150 km away. The victim testified that the offender had a foreign accent. According to the database details, suspect 2 could fit this criterion.

Samples from both suspects and the male fetus were analyzed using the AmpFℓSTR® YFiler® and subsequently searched against an internal Y-haplotype database (400 men) and against the ABI YFiler haplotype database (11,393 men). The suspects, the victim and the fetus were subjected to additional autosomal short tandem repeat (STR) analysis, using the AmpFℓSTR® Minifiler™ and PowerPlex® 16 kits (Promega Corp., Madison WI), analyzing 17 STR loci in total and the amelogenin locus.

In this article, the possibility of searching a database for obligatory alleles in a fatherless paternity case is presented, and the statistical analysis of the results and implications of such a procedure for investigative purposes is discussed.

Materials and Methods

Sampling and DNA Profiling

Four samples of approximately 2 mm^3 were sampled from fetal tissue. Reference samples were sampled from FTA cards.

DNA was extracted by the Chelex procedure (1) in a final volume of 200 μL. The DNA yields in all samples were estimated using the Real Time PCR Quantifiler® Duo DNA Quantification Kit (Applied Biosystems) (2) and the ABI PRISM® 7500 instrument (Applied Biosystems). Amplification reactions were performed using the following kits: the AmpFℓSTR® SGM Plus®, the AmpFℓSTR® MiniFiler™, the AmpFℓSTR® YFiler®, and PowerPlex® 16 according to the manufacturers' recommendations (3–6).

[1]Forensic DNA and Biology Laboratory, Division of Identification and Forensic Science (DIFS), Israel Police National Headquarters, Jerusalem, Israel.
[2]DNA Database Laboratory, Division of Identification and Forensic Science (DIFS), Israel Police National Headquarters, Jerusalem, Israel.
[3]Department of Statistics and Department of Ecology, Evolution and Behavior, The Hebrew University of Jerusalem, Jerusalem, Israel.
[*]Present address: Faculty of Health, Science and Medicine, Bond University, Gold Coast, QLD 4229, Australia.

Received 7 Dec. 2010; and in revised form 4 June 2011; accepted 18 June 2011.

The amplified PCR products were separated by electrophoresis on an ABI PRISM® 3130*xl* Genetic Analyzer fitted with a 50 μm by 36-cm capillary, loaded with POP-4 polymer. The running conditions were as follows: 5 sec injection time, 30 min run time, and 60°C running temperature. GeneMapper® 3.2 software (Applied Biosystems) was used for genotyping. An analytical threshold of 60 relative fluorescence units (RFU) was used for allele labeling, and a stochastic threshold of 200 RFU was set for designation of homozygotes.

Database Search for Obligatory Alleles

The database was searched for the seven obligatory parental alleles (underlined alleles in Table 1). In the other three loci, the mother and fetus were identical heterozygotes, resulting in two possible paternal alleles. Our database search program does not allow searching several possible allele combinations; therefore, the search was conducted using obligatory alleles only. The Israeli National Reference Database contained 55,000 individuals at the time of the search. This database contains reference samples of suspects, convicted individuals, and prisoners, all suspected or convicted in sexual, violence, and volume crimes. The Y-haplotype search was done using an internal database of 400 men (of Jewish and Arab origin) and an Applied Biosystems YFiler haplotype database, consisting of 11,393 men (7).

Statistical Calculations

For each autosomal marker, the paternal allele of the fetus (or the two possible alleles, if the fetus and the mother were both identical heterozygote in that marker) was identified, and the population frequency of the carriers of these alleles was estimated. These estimates are based on an internal Israeli police database, which consists of 300 individuals from each of the Jewish and the Arab ethnic groups, comprising together the vast majority of the country's population. The presented results are the weighted average of these two ethnic groups, based on their relative size in the Israeli population. There is no further subcategorization of other subgroups in these populations. The same database served to calculate the paternity index (PI) for each suspect, which is the ratio of the probability that a man with the suspect's type would produce a child of the observed type to the probability that a random person from the population could produce a child of the observed type.

As an estimate of the population frequency p of a Y-haplotype that is not present in a database of size N, one can use $\hat{p} = \frac{2}{N+2}$ (8) or alternatively use $\hat{p} = 1 - \sqrt[N]{\alpha}$ (9). The latter is, in fact, the upper confidence limit of an exact, one-sided 1-α confidence interval for the population frequency p of that haplotype. In other words, it sets the probability of not observing that haplotype in the police database [i.e., $(1 - p)^N$] to be at least a prespecified, small value α. Choosing α to be the commonly accepted value of 0.05 produces a more conservative estimate of p than the one suggested by Balding et al. (8). In fact, $\hat{p} = 1 - \sqrt[N]{\alpha}$ is more conservative than $\hat{p} = \frac{2}{N+2}$ for every $\alpha \le e^{-2}$.

Results and Discussion

Summary of the Performed Tests

Of the four samples obtained from the fetal tissue, one sample revealed a full AmpFℓSTR® SGM Plus® male profile (Table 1). This sample was subsequently amplified using the following kits: the AmpFℓSTR® Y-Filer™, AmpFℓSTR® MiniFiler™, and PowerPlex®

TABLE 1—*Summary of the results of the autosomal STR analysis of the victim, her aborted fetus, and two probable suspects that could not be excluded, following the database search.*

	Loci																			
	AmpFℓSTR® SGM Plus® Kit											Additional Loci from AmpFℓSTR® Minifiler™ Kit			Additional Loci from PowerPlex® 16 Kit					
Sample	FGA	THO1	D19S433	D18S51	D21S11	D8S1179	Amelogenin	D2S1338	D16S539	vWA	D3S1358	D7S820	CSF1PO	D13S317	D13S317	Penta E	D5S818	Penta D	TPOX
Fetus	21,26	7,7	15,15.2	12,16	30,34.2	13,14	XY	20,21	8,10	17,19	16,17	10,11	11,12	**12**	**9,12**	12,18	11,12	9,12	8,10
Victim	20,26	7,9	13,15	12,17	29,34.2	13,14	XX	20,21	8,11	16,17	16,17	11,11	11,12	**8**	**8,9**	12,21	11,12	9,12	9,10
Suspect 1	21,24	6,7	14,15.2	16,17	30,30	13,13	XY	17,20	10,13	16,19	16,17	9,10	12,12	11,12	11,12	12,18	11,13	9,9	8,9
Suspect 2	21,22	7,9.3	13,15.2	16,20	30,30	13,16	XY	16,20	10,11	17,19	17,19	10,12	11,12	12,12	12,12	12,12	13,13	9,10	11,11

Paternal obligatory alleles are underlined. Note the difference in the genotypes of the victim and the fetus in D13S317 marker between Minifiler™ and PowerPlex® kits (shown in bold).

16. The reference samples of the victim and the alleged fathers were amplified accordingly. The remaining three fetal samples showed a partial profile using the AmpFℓSTR® SGM Plus®, which was consistent with the obtained full profile (data not shown). The analysis of these samples was not continued.

Database Search for Obligatory Alleles

The search for the seven obligatory alleles in the reference DNA database (see underlined alleles in Table 1) revealed three individuals that could not be excluded as potential fathers of the fetus. Further comparison of the full SGM Plus kit (10 STR markers) of the fetus, victim, and the three suspects identified by the database search excluded "suspect 3" as a probable father. The PIs of the remaining two possible biologic fathers were 2,039,000 for suspect 1 and 510,000 for suspect 2.

Database Search for the Obtained Y-Haplotype

Because of the male origin of the fetus, Y chromosome profiling was performed. The analysis of the Y-Filer (Table 2) did not result in exclusion of suspect 1 as a possible father of the tested fetus. Suspect 2, on the other hand, was excluded in this test. The Y-haplotype of suspect 1 matched the Y-haplotype of the fetus. This Y-haplotype existed neither in the internal database (400 men) nor in the Applied Biosystems YFiler haplotype database (11,393 men).

Combining the Results of Three STR Kits

Simultaneously, additional autosomal STR markers genotyping was performed. Analysis of three additional autosomal STR loci, (CSF1PO, D7S820, and D13S317) using the Minifiler™ kit did not exclude either suspect 1 or suspect 2 (Table 1) for all 13 loci tested.

Analysis of the D13S317 locus, present in the AmpFℓSTR® Minifiler™ kit, indicated an inconsistency between the type of the mother and the fetus. The mother's type was 8, and the fetus' type was 12 (Table 1, alleles emphasized in bold). Further examination of this locus revealed that the height of the fluorescent intensity peaks in both samples, in comparison with adjacent loci, was lower than expected from a homozygote profile (data not shown). These data suggested the existence of a null allele in the profiles of the mother and the fetus, as has been previously described in the D13S317 locus (10,11).

The genotyping by PowerPlex® 16 kit indicated that the actual profile of the mother at locus D13S317 was 8,9, and the profile of the fetus was 9,12 (Table 1, shown in bold); thus, this locus as well did not exclude either suspect 1 or suspect 2. Only the testing of four additional loci (Penta E, D5S818, Penta D, and TPOX) enabled the exclusion (based on three of these loci) of suspect 2 as a possible father of the fetus. Suspect 1 was not excluded in all 17 autosomal STR loci tested. These results were consistent with the Y haplotyping results (Table 2).

Statistical Analysis

The probability that a random male in the Israeli population will not be excluded as a potential father of the fetus as represented in Table 1 is 1 in 455,000, when using the routinely applied SGM Plus kit (10 markers).

The National Reference Database contained 55,000 individuals at the time of the search. Thus, if it is assumed that the biologic father is among these 55,000 profiles, the probability of finding at least one more individual in the database that could not be excluded as a potential father of the fetus is approximately $1 - \exp(-\frac{55{,}000}{455{,}000}) = 11.4\%$ (using the Poisson approximation), and this is not an insignificant probability at all.

Combining these results with the AmpFℓSTR® Minifiler™ kit, the probability of nonexclusion decreases to 1 in 1,142,000 in the Israeli population. Subsequently, the probability of finding at least one more individual in the database that could not be excluded as a potential father of the fetus decreases to 4.7%.

Adding the PowerPlex® 16 kit, the probability of nonexclusion decreases to 1 in 87.8 million in the Israeli population. The probability of finding at least one more individual in the database that could not be excluded decreases to 0.063%.

Based on the internal police Y-haplotype database of 400 men, the estimated population frequency of the fetus' Y-haplotype is $1 - 0.05^{(1/400)} = 0.00746$, that is, 1 in 134. For comparison, the estimated frequency is only 0.00026 when using the larger, 11,393 males ABI database.

Evaluation of the Results and Potential Recommendations

Cold hit statistics have been extensively discussed in the literature following the database similarity search and performed in Arizona, and different methodologies have been suggested (12–16). These results indicate the need to extend the statistical evaluation and discuss further issues. The frequency of a specific DNA profile remains as calculated, yet the probability of finding a matching profile in the database should be recalculated, taking under consideration the size of a database.

In light of this information, searching a database only for obligatory alleles may become problematic. It should also be kept in mind that the search for obligatory alleles in a database is in a manner similar to familial searching. Familial searches in databases have been extensively debated, and in such cases, additional markers are highly recommended (12–18).

The individuals who are not excluded as a possible father in a sexual assault case should be retested for more independent autosomal STR loci, and if the fetus is of male origin, haplotyping of chromosome Y is essential.

TABLE 2—*Summary of the analysis of the Y chromosome markers (AmpFℓSTR® YFiler®) of the aborted fetus and the probable suspects that could not be a priori excluded.*

	Loci															
Sample	DYS448	DYS438	DYS437	GATA_H4	DYS392	DYS635	DYS439	DYS391	DYS393	DYS385	DYS19	DYS458	DYS389II	DYS390	DYS389I	DYS456
Fetus	20	10	16	11	11	22	11	11	13	13,14	14	15	29	22	12	15
Suspect 1	20	10	16	11	11	22	11	11	13	13,14	14	15	29	22	12	15
Suspect 2	20	10	14	12	11	21	10	9	13	14,15	13	17	30	24	14	15

In this particular case, both suspects resulted in relatively high PIs using the AmpFℓSTR® SGM Plus®. One should consider a scenario in which suspect 1 was not in the database. In such a case, only suspect 2's would be identified as the possible perpetrator, and the statistical weight of this evidence may look substantial if the database size is not considered, especially taking into account the victim's testimony, which matched suspect 2's ethnical description. In this case, 13 autosomal STR loci were inadequate to distinguish between two suspects as being possible fathers of the tested fetus. Of 17 autosomal loci tested, suspect 2 was excluded at three loci only.

Based on these extensive results of 17 autosomal loci, the PI for suspect 1 was calculated to be 261 million to 1. The Y-haplotype analysis, with its PI of 134, yielded an additional confirmation to the involvement of suspect 1.

It is recommended that the search of obligatory alleles in a database should be treated as an investigative tool and therefore should be supported by more evidence. This will require the investigating unit to follow up the DNA lead and carry on a thorough investigation to collect more information that may enlighten the details of the investigated case. In this case, following the DNA analysis and before approaching the suspect, the victim was requisitioned. Upon questioning, the investigator brought up the name of suspect 1 and requested to know what kind of relations the victim had with this suspect. The victim admitted that she had had an affair with this man. As a result of the affair, she became pregnant and invented the rape scenario. She was afraid of the reaction of her family and boyfriend who actually suspected her. Suspect 1, who was not excluded as the father of the fetus, was not at any time involved in a sexual assault of any kind.

Our case demonstrates that a man in a database who possesses obligatory paternal alleles may or may not be the perpetrator of a rape and father of a child of rape. If a database is large or few marker loci were tested, innocent individuals could become suspects. Even when many autosomal and Y chromosomal loci are evaluated, the actual perpetrator may be a patrilineal relative of the man found in the database.

Conclusion

The search of databases in rape paternity cases, as in other criminal cases, should be approached with extreme caution. The search may provide false positives as we have shown. A "hit" should be regarded only as a preliminary finding and must be supported by matches at a greater number or more informative loci as well as by nongenetic evidence.

References

1. Walsh PS, Metzger DA, Higuchi R. Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 1991;10:506–13.
2. Quantifiler® kit user's manual. The Perkin-Elmer Corporation. Foster City, CA: Applied Biosystems, 2003.
3. AmpFℓSTR® SGM Plus® PCR Amplification Kit, user's manual. Foster City, CA: Applied Biosystems, 2006, http://www3.appliedbiosystems.com/cms/groups/applied_markets_support/documents/generaldocuments/cms_041049.pdf (accessed December 21, 2009).
4. AmpFℓSTR® MiniFiler™ PCR Amplification Kit, user guide. Foster City, CA: Applied Biosystems, 2007, http://www3.appliedbiosystems.com/cms/groups/applied_markets_support/documents/generaldocuments/cms_042748.pdf (accessed December 21, 2009).
5. AmpFℓSTR® YFiler® Amplification Kit, user's manual. Foster City, CA: Applied BioSystems, 2006, http://www3.appliedbiosystems.com/cms/groups/applied_markets_support/documents/generaldocuments/cms_041477.pdf (accessed December 21, 2009).
6. PowerPlex® 16 System, technical manual. Madison, WI: Promega Corporation, 2008, http://www.promega.com/tbs/tmd012/tmd012.pdf (accessed December 21, 2009)
7. YFiler Haplotype Database, http://www6.appliedbiosystems.com/yfiler-database/(accessed December 21, 2009).
8. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 1994;64:125–40.
9. Budowle B, Ge J, Chakraborty R. Basic principles for estimating the rarity of Y-STR haplotypes derived from forensic evidence. Proceedings of the Eighteenth International Symposium on Human Identification; 2007 Oct 1–4; Hollywood, CA. Madison, WI: Promega Corporation, 2007.
10. Boutrand L, Egyaed B, Füredi S, Mommers N, Mertens G, Vandenberghe A. Variations in primer sequences are the origin of allele drop-out at loci D13S317 and CD4. Int J Legal Med 2001;114:295–7.
11. Hill CR, Kline MC, Mulero JJ, Lagacé RE, Chang CW, Hennessy LK, et al. Concordance study between the AmpFℓSTR MiniFiler PCR amplification kit and conventional STR typing kits. J Forensic Sci 2007;52(4):870–3.
12. Brenner C. Arizona DNA database matches, 2007, http://dna-view.com/ArizonaMatch.htm (accessed December 21, 2009).
13. Devlin K. Scientific heat about cold hits, 2007, http://www.stanford.edu/~kdevlin/Cold_Hit_Probabilities.pdf (accessed December 21, 2009).
14. Budowle B, Planz JV, Chakraborty R, Callaghan TF, Eisenberg AJ. Clarification of statistical issues related to the operation of CODIS. Proceedings of the Seventeenth International Symposium on Human Identification; 2006 Oct 10–12; Nashville, TN. Madison, WI: Promega Corporation, 2006.
15. Budowle B, Baechtel FS, Chakraborty R. Partial matches in heterogeneous offender databases do not call into question the validity of random match probability calculation. Int J Legal Med 2009;123:59–63.
16. Song YS, Patil A, Murphy EE, Slatkin M. Average probability that a "cold hit" in a DNA database search results in an erroneous attribution. J Forensic Sci 2009;54(1):22–7.
17. Bieber FR, Brenner CH, Lazer D. Finding criminals through DNA of their relatives. Science 2006;312:1315–6.
18. Wenk RE, Chiafari FA, Gorlin J, Polesky HF. Better tolls are needed for parentage and kinship studies. Transfusion 2003;43:979–81.

Additional information and reprint requests:
Mark Barash, M.Sc.
Faculty of Health, Science and Medicine
Bond University
Gold Coast
QLD 4229
Australia
E-mail: mbarash@bond.edu.au