Forensic Science International: Genetics Supplement Series 5 (2015) e50–e51

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

The open-source software LRmix can be used to analyse SNP mixtures

Peter Gill[a,b,*], Hinda Haned[c], Mayra Eduardoff[d], Carla Santos[e], Chris Phillips[e], Walther Parson[d,f]

[a] Norwegian Institute of Public Health, Oslo, Norway
[b] Department of Forensic Medicine, University of Oslo, Norway
[c] Department of Human Biological Traces, Netherlands Forensic Institute, The Hague, The Netherlands
[d] Institute of Legal Medicine, Medical University of Innsbruck, Austria
[e] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Spain
[f] Forensic Science Program, The Pennsylvania State University, PA, USA

ARTICLE INFO

Article history:
Received 14 August 2015
Accepted 7 September 2015
Available online 21 September 2015

Keywords:
SNP
LRmix
Complex mixture
Likelihood ratio

ABSTRACT

A series of two- and three-person mixtures of varying ratios were prepared and analysed with Life Technologies' HID-Ion AmpliSeq™ Identity Panel v2.2 using the Ion PGM™ massively parallel sequencing system. This panel includes 136 autosomal SNPs and 33 Y-chromosome SNPs. Using the reference samples of the mixture donors, we evaluated the strength of evidence with likelihood ratio (LR) calculations using the open-source LRmix program. This program was designed for multi-allelic STRs, but can be extended to bi-allelic SNPs without modification. We conditioned on each donor in turn, as the person of interest (POI) in the numerator. The LR tests showed that two-person mixtures typically gave LRs of the order of $>10^8$ whereas three-person mixtures were highly variable in results. Approximately 50% of results were exclusionary or neutral. The remainder provided LRs ranging between 10^3–10^8. We further evaluated these mixtures using diagnostic non-contributor tests. Our preliminary work shows that simple two-person mixtures can be readily analysed with LRmix, but the performance of three- or more person mixtures is less predictable and may fail to provide probative evidence. However, if a higher number of loci were multiplexed, the analysis of mixtures would be improved, particularly if less frequent variation (allele frequencies below 0.1) was targeted in identity SNP selection.

1. Introduction

The study described is the first demonstration to indicate SNP genotypes of a set of mixed-source samples, analysed with Life Technologies' HID-Ion AmpliSeq™ Identity Panel v2.2 and using the Ion PGM™ massively parallel sequencing system [1], can be reliably interpreted using the complex mixture interpretation software LRmix [2].

The program LRmix follows the recommendations of the ISFG DNA commission on the interpretation of STR mixtures [3]. SNP mixtures can be evaluated in the same way since SNPs behave in the same way as two-allele STRs, albeit with many more loci to consider, but with the advantage that there are no stutters to consider.

2. Materials and methods

Mixtures were prepared from three individuals, labelled CT, MG and CQ. A series of mixtures were prepared in the proportions indicated in Table 1.

Two person mixtures comprised the ratios of 1:1, 1:3 and 1:9 plus reverse ratios of 3:1, 9:1. For each possible combination of contributors this provided a total of ten likelihood ratios to analyse. For three person mixtures, the ratios utilised were 1:1:1, 1:1:5, 1:5:1, 1:5:5, 5:1:1, 5:1:5 and 5:5:1. For three contributor arrangements per mixture ratio, there were a total of 21 likelihood ratios to analyse. All mixture ratios were analysed in duplicate.

The likelihood ratios tested were:

$$\text{Two} - \text{person mixture}: \quad \text{LR} = \frac{\text{S, U}|Hp}{\text{U, U}|Hd}.$$

$$\text{Three} - \text{person mixture}: \quad \text{LR} = \frac{\text{S, U, U}|Hp}{\text{U, U, U}|Hd}.$$

* Corresponding author at: Norwegian Institute of Public Health, Oslo, Norway.
E-mail address: peterd.gill@gmail.com (P. Gill).

http://dx.doi.org/10.1016/j.fsigss.2015.09.020
1875-1768/

Table 1
A list of mixtures prepared from three different individuals (CQ, MG and CT) in ratios indicated in the left hand column. LRs ($\log_{10}$) are provided, along with non-contributor statistics summarised from one million simulations of random man.

			Results of 1 million non contributor tests						
Mixture	Proposition	LR (log)	Min	0.01	0.05	0.5	0.95	0.99	Max
1MG:1CT	CT,U	9.597	−59	−42	−37	−26	−15	−11	4.5
1:1	MG,U	8.374	−100	−73	−65	−46	−29	−21	−1.3
1_3	CT,U	9.597	−59	−42	−38	−26	−15	−11	3.02
1_3	MG,U	8.377	−60	−42	−38	−26	−15	−11	2.3
1_9	CT,U	9.597	−59	−42	−38	−26	−15	−11	2.5
1_9	MG,U	8.377	−59	−42	−38	−26	−15	−11	2.8
3_1	CT,U	9.597	−59	−42	−38	−26	−15	−11	2.09
3_1	MG,U	8.377	−59	−42	−38	−26	−15	−11	3.23
9_1	CT,U	1.342	−33	−32	−30	−22	−14	−11	−10
9_1	MG,U	14.01	−46	−34	−30	−22	−13	−10	2.3
1MG:1CQ:1CT	CT,U,U	2.11	−32	−23	−19	−10.5	−2.4	1.3	3.16
1:1:1	MG,U,U	1.76	−32	−23	−19	−10.5	−2.4	−1.3	3.18
1:1:1	CQ,U,U	1.62	−16.5	−11	−9	−4.5	−0.34	1.2	3.9
1:1:5	CT,U,U	−4.5	−20	−13	−11	−6.5	−2	−0.51	5.06
1:1:5	MG,U,U	8.159	−20	−12	−11	−6.6	−2.5	−0.88	4.27
1:1:5	CQ,U,U	5.269	−30	−20	−17	−11	−5.3	−3	3.94
1:5:1	CT,U,U	−7.715	−27	−19	−16	−10	−4	−1.7	5.3
1:5:1	MG,U,U	6.077	−21	−13	−11	−7	−2.4	−0.68	4.35
1:5:1	CQ,U,U	6.033	−21	−13	−11	−7	−2.4	−0.7	4.91
1:5:5	CT,U,U	4.197	−20	−13	−11	−6	−1.7	−0.01	4.15
1:5:5	MG,U,U	−7.722	−50	−34	−29	−18	−7.5	−3.4	4.3
1:5:5	CQ,U,U	3.505	−20	−13	−11	−6	−1.7	0.01	4.2
5:1:1	CT,U,U	2.132	−32	−23	−19	−11	−2.4	1.3	3.3
5:1:1	MG,U,U	1.764	−32	−23	−19	−10.5	−2.4	1.3	3.1
5:1:1	CQ,U,U	1.47	−20	−14	−11	−6	−0.8	1.26	3.45
5:1:5	CT,U,U	4.798	−15.5	−10	−8.5	−5	−1	0.33	3.9
5:1:5	MG,U,U	4.33	−19	−12	−10	−5.5	−1.4	0.11	4.9
5:1:5	CQ,U,U	−14.28	−57	−39	−34	−21	−9	−4.5	4.7
5:5:1	CT,U,U	−27	−48	−34	−30	−20	−11	−8	3.16
5:5:1	MG,U,U	8.34	−27	−18	−16	−10.5	−5	−3	4.6
5:5:1	CQ,U,U	7.829	−22	−16	−14	−8.5	−4	−2	4.3

Where S is the 'suspect', comprising either CT, MG, or CQ in turn, and U is an unknown (unrelated) individual conditioned under the prosecution hypothesis (Hp) or the alternative defence hypothesis (Hd).

LRmix was used without any modification, using European allele frequencies as described by [4] and Fst = 0.01; Pr (drop-in) = 0.05. Non contributor tests (1 million) were carried out as described by Gill et al. [5]. These tests substitute a simulated random man for the 'suspect', designated as 'S' in the numerator of the likelihood ratio. We naturally expect that a true contributor should provide a high LR, whereas a false contributor (provided by a random man) should give a much lower LR.

3. Results and discussion

For a complete review of the software used to interpret mixtures see Gill et al. [6].

Amongst the observations made using LRmix to interpret SNP genotypes, the increased number of loci did not affect computation speeds. Two person mixtures gave LRs > $\log_{10}(8)$ with one exception (CT,U) where the LR = $\log_{10}(1.3)$. However this was a 9:1 (U:CT) mixture, where there was significant dropout of CT alleles in this sample. Three person mixtures proved to be much more variable. Several LRs were less than 1 (exclusionary), and not all were coincident with mixtures where the conditioned individual in the numerator was a minor contributor. Apart from allelic and locus dropout, the LR may be substantially reduced by the effect of 'allelic saturation' which is an inevitable consequence of increasing the number of contributors to the mixture. This results in heterozygote (*ab*) crime-stain profiles matching all reference individuals in a population (*aa, ab* or *bb*) hence the strength of the evidence tends towards neutrality. This is amply illustrated by the 1:1:1 mixtures in Table 1 where all LRs recorded were very low.

Nevertheless, approximately 50% of three person profiles gave LRs > $\log_{10}(4)$ with a maximum observed of $\log_{10}(8)$. It is also clear from the non-contributor tests that samples providing high LRs gave correspondingly low non-contributor distributions, which indicates that a LR which provides a probative strength of evidence is reliable. We are now carrying out research to discover if the 'coverage' per allele (analogous to peak height in STRs) provides additional useful information.

Funding

PG, CP, ME and WP have received funding support from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 285487. CS was supported by grant SFRH/BD/75627/2010 from the Portugese Foundation for Science and Technology (FCT).

References

[1] http://www.lifetechnologies.com/content/dam/LifeTech/global/applied-sciences/pdfs/human-identification/CO119694-HID-Ion-AmpliSeq-Identity-panel-Flyer-Americas-FLR.pdf.
[2] H., Haned, LRmix Tutorial, version 4.1 http://forensim.r-forge.r-project.org/misc/LRmix.pdf.
[3] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures, Forensic Sci. Int. 160 (2–3) (2006) 90–101.
[4] J. Amigo, A. Salas, C. Phillips, ENGINES: exploring single nucleotide variation in entire human genomes, BMC Bioinf. 12 (1) (2011) 105.
[5] P. Gill, P.H. Haned, A new methodological framework to interpret complex DNA profiles using likelihood ratios, Forensic Sci. Int. Genet. 7 (2) (2013) 251–263.
[6] P. Gill, H. Haned, O. Bleka, O. Hansson, G. Dørum, T. Egeland, Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches—twenty years of research and development, Forensic Sci. Int. Genet. (2015) in press.