



Automated estimation of the number of contributors in autosomal STR profiles

Corina Benschop*, Anouk Backx, Titia Sijen

Division of Biological Traces (R&D), Netherlands Forensic Institute, the Netherlands



ARTICLE INFO

Keywords:

Forensic science
PowerPlex® Fusion 6C
Number of contributors
Maximum allele count
Total allele count
nC-tool

ABSTRACT

Estimating the number of contributors to mixed STR profiles can be complex. This study describes the nC-tool to assist DNA expert in this process. The nC-tool is based on the total allele count for PowerPlex® Fusion 6C profiles and showed improved performance when compared to the maximum allele count approach.

1. Introduction

Estimating the number of contributors (NOC) is part of the interpretation process of autosomal STR profiles. This can be important, *e.g.* when defining hypotheses for calculating the weight of evidence. Estimating the NOC can be complicated with a large NOC, high allele sharing, stochastic effects and/or degraded DNA. The most commonly and easily used method to estimate the NOC is the maximum allele count (MAC) approach that regards the locus with the largest number of alleles which is divided by two (rounded up) and provides an indication for the minimum NOC. Besides the MAC, the total allele count (TAC) is one of the profile characteristics that can be informative for the NOC. In this study, a number of contributors tool, nC-tool, was developed that is based on simulations for the TAC.

2. Materials and methods

A Dutch allele frequencies database [1] was used to simulate 1p to 5p PowerPlex® Fusion 6C (PPF6C) profiles with drop-out ranging from 0% to 50%, in steps of 5%. For each step and each NOC 10,000 profiles were simulated in R (based on a method described in [2]) and for each profile the TAC of the autosomal markers was noted. Next, for each TAC the probabilities were calculated per NOC (1, 2, 3, 4, 5) in four categories of drop-out (0%, 1–10%, 11–25%, 26–50%). The simulated DNA profiles did not exhibit artefacts, such as elevated stutter or drop-in. In real casework these artefacts do occur, especially with low-template amounts of DNA, and can increase the TAC. Therefore, the TACs obtained from laboratory-generated profiles were corrected by lowering

the number of alleles by 5% (the estimated % of drop-in/elevated stutter per profile) prior to comparison to the simulated TACs. This approach was programmed in an Excel spreadsheet to which the TAC from PPF6C profiles can be entered after which a bar graph shows the probabilities per NOC and drop-out range (Fig. 1).

Laboratory-generated 2p-5p PPF6C profiles from [3] were used to test how well the corrected TACs of simulated PPF6C profiles are informative on the NOC. These profiles represent a variety of mixtures with variation for genotypes ($n = 30$ different donors), mixture proportion, drop-out and allele sharing. The TAC for each individual replicate as well as the TAC for the composite of three replicates combined was entered in the nC-tool where they are corrected by 5% and the NOC is inferred. Profiles resulting a 5% corrected TAC larger than 143 ($n = 2$) were omitted from the dataset as the nC-tool covers the TAC ranging from 8-143. The NOC that gave the highest probability when using the true drop-out rate was noted and compared to the NOC inferred from the MAC approach. An estimate of five contributors using the nC-tool was interpreted as a minimum of five contributors as the tool does not provide a larger number than five. The tool is available on request.

3. Results and discussion

Mischaracterization rates were lower with the TAC nC-tool than with the MAC method (Table 1). The percentage of profiles with a correctly estimated NOC improved by 15% and 42% for individual profiles or composite profiles derived of three replicates, respectively. Improved performance was observed for each NOC category (Table 1).

* Corresponding author at: Laan van Ypenburg 6, The Hague, 2497 GB, the Netherlands.
E-mail address: c.benschop@nfi.nl (C. Benschop).

Type TAC here (total number of alleles at the autosomal PPF6C loci):

102

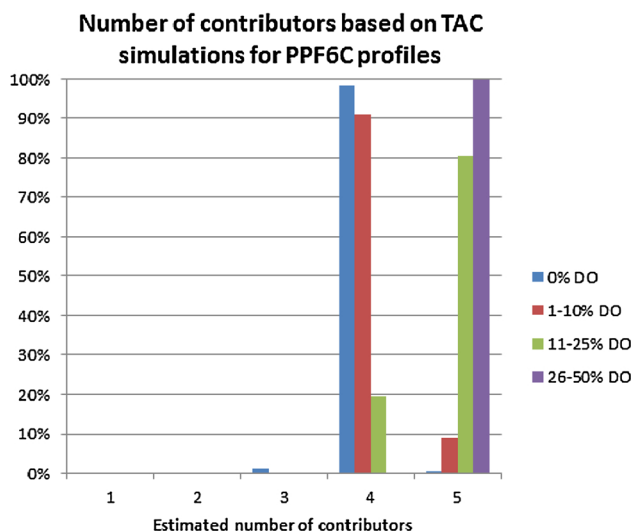


Fig. 1. Illustrative nC-tool result for a PPF6C profile with TAC = 102. The nC-tool corrects the TAC by 5% for elevated stutter/drop-in and graphically presents the estimated NOC per drop-out (DO) category.

Table 1

Performance of the TAC nC-tool and the MAC-method for inference of the NOC to mixed PPF6C DNA profiles. Results for A) individual replicates and B) composite profiles derived of three replicates.

A	Number of profiles per drop-out category				TAC nC-tool			MAC		
	0%	1–10%	11–25%	26–50%	Under-estimated	Correctly estimated	Over-estimated	Under-estimated	Correctly estimated	Over-estimated
True number of contributors (n per MAC/TAC method)										
2 (n = 90)	41	33	16	0	0%	100%	0%	0%	68%	32%
3 (n = 88)	35	39	12	2	0%	89%	11%	1%	85%	14%
4 (n = 89)	30	43	16	0	12%	74%	30%	17%	71%	12%
5 (n = 87)	27	42	17	1	36%	64%	not applicable	47%	46%	7%
Total (n = 354)	133	157	61	3	12%	82%	10%	16%	67%	16%
B	Number of profiles per drop-out category				TAC nC-tool			MAC		
	0%	1–10%	11–25%	26–50%	Under-estimated	Correctly estimated	Over-estimated	Under-estimated	Correctly estimated	Over-estimated
True number of contributors (n per MAC/TAC method)										
2 (n = 30)	20	9	1	0	0%	97%	17%	0%	33%	67%
3 (n = 29)	19	9	0	0	0%	93%	21%	0%	33%	67%
4 (n = 29)	19	9	1	0	9%	66%	49%	3%	55%	41%
5 (n = 25)	15	8	2	0	17%	80%	not applicable	28%	48%	24%
Total (n = 112)	73	35	4	0	6%	84%	22%	8%	42%	50%

With 5p mixtures, however, the high percentages of correct estimates are likely to be overrepresented as overestimating is not possible with the nC-tool. Underestimations occurred least frequent and were observed mostly with mixed profiles for which the donors shared many alleles and/or exhibited many homozygous loci. Overestimations were observed more often, and were obtained even with the MAC method. These occurred predominantly for mixed profiles with low allele sharing between contributors and/or low-template profiles exhibiting elevated stutter/drop-in. Individual replicates and composite profiles of three replicates showed similar trends with regard to under- and overestimations (Table 1).

The nC-tool showed an overall better performance, i.e. less mischaracterizations, for the 2p-5p PPF6C profiles when compared to the MAC method. Hence, the TAC of the autosomal markers of PPF6C was more informative on the NOC when compared to the MAC approach

with the profiles used in this study. In actual casework mischaracterization rates can be higher as we used the true drop-out rate which is unknown in real cases. Furthermore, using more characteristics such as peak height information and allele frequencies, and combining these in a tool for inference of the NOC, might result in even less mischaracterizations and were examined in another study using a machine learning approach (Benschop et al., manuscript in preparation).

4. Concluding remarks

The TAC nC-tool can be a useful tool for inference of the NOC to mixed autosomal PPF6C profiles, additionally to the MAC method and profile examination. The applicable domain of the nC-tool is PPF6C profiles with a corrected TAC of 8–143, having up to five contributors and with a maximum of 50% drop-out. The nC-tool can be applied to the TAC of individual replicates as well as for composite profiles. Note that the tool is developed for PPF6C profiles; we do not recommend application to other STR profiling systems even when comprising the exact same loci.

Declaration of Competing Interest

None.

Acknowledgement

We are thankful to Øyvind Bleka for providing the initial R script to simulate mixed profiles and record the TACs.

References

- [1] A.A. Westen, T. Kraaijenbrink, E.A. Robles de Medina, et al., Comparing six commercial autosomal STR kits in a large Dutch population sample, *Forensic Sci. Int. Genet.* 10 (2014) 55–63.
- [2] Ø. Bleka, C.C.G. Benschop, G. Storvik, P. Gill, A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles, *Forensic Sci. Int. Genet.* 25 (2016) 85–96.
- [3] C.C.G. Benschop, A. Nijveld, F.E. Duijs, T. Sijen, An assessment of the performance of the probabilistic genotyping software EuroForMix: trends in likelihood ratios and analysis of type I & II errors, *Forensic Sci. Int. Genet.* 42 (2019) 31–38.