Experiments on the number of contributors, two fewer and two more.

Kevin Cheng, Jo-Anne Bright, and John Buckleton

Introduction

In casework, the true number of contributors (NoC) to a questioned profile is always unknown. Analysts are likely to add contributors in the presence of ambiguous peaks such as artefact peaks or inflated stutter peaks. The assumption of one fewer contributor may be made when contributors are at very low levels and dropping out, in profiles where DNA is from individuals with similar profiles, or in family scenarios where mixtures of related individuals might reasonably be expected to be recovered.

The effect of the uncertainty in the number of contributors within STRmix™ has previously been reported for a number of profiles with N, N-1, and N+1 assumed contributors, where N is the number of contributors [1-5]. The under assignment of NoC (N-1) had the effect of assigning exclusionary *LR*s ($LR = 0$) to one or more contributors within the profiles. Whereas, the inclusion of an additional contributor (N+1) beyond that present in the profile had the effect of lowering the *LR* for true trace contributors within the profile. STRmix™ adds the additional (unseen) profile at trace levels which interacts with the known trace contribution, diffusing the genotype weights and lowering the *LR*. There was no significant effect on the *LR* of the major or minor contributor within the profiles. N+1 had the effect of raising some non-donors from exclusionary *LR*s to *LR*s near 1.

It is possible for a mixture to be interpreted assuming two fewer or two additional contributors than the target N. However, not much work has been done to show the impact of assuming N-2 or N+2 contributors to a mixture. Therefore, the impact of assuming N-2 or N+2 contributors on the *LR* was further investigated for 49 GlobalFiler™ mixtures from the PROVEDIt dataset [6].

Methods

Each of the 49 mixtures was interpreted using the experimental design number of contributors (N), assuming two fewer contributors than N (N-2), and assuming two additional contributors than N (N+2). The results were compared to a database containing 223 individuals using the Database Search function. The database contained reference profiles for known contributors to the mixtures and profiles of known non-contributors artificially generated using the FBI Extended Caucasian allele frequency database.

Using the FBI Extended Caucasian allele frequencies and an F_{ST} of 0.01, a sub-source point estimate *LR* was assigned, where the propositions considered were:

H_p: The DNA originated from the database individual and N-1 unknown unrelated individuals

H_a: The DNA originated from N unknown unrelated individuals

Where N is the assumed number of contributors to the profile, this could be the experimental design number of contributors (N), assuming two fewer contributors than N (N-2), or assuming two addition contributors than N (N+2).

Only 30 of the 49 mixtures were interpreted assuming N-2 (where N-2=0 these profiles were not interpreted). Of these 30, only 5 were able to be explained assuming N-2 and therefore an *LR* was able to be assigned.

Results

Figure 1: Comparison of log(*LR*) obtained when profiles were run in STRmix™ as N versus N-2 contributors for the known and non-contributors (where H_p true is represented by blue circle data points and H_a true are represented by red circles).

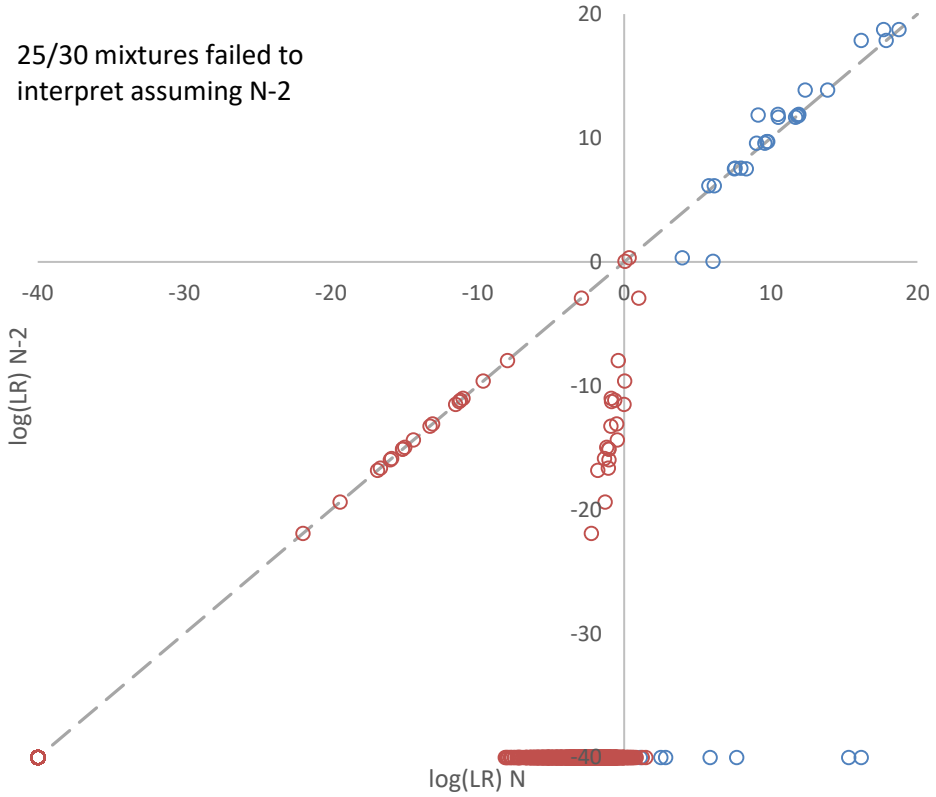

Figure 1 demonstrates that, in general, there is no significant effect on the *LR* of the 'true' major contributors to the mixture if the number of contributors is underestimated by two fewer contributors to the mixture. This is demonstrated by the large *LR*s (>log(*LR*)~8) remaining on the $x = y$ line that are. We do note that some data have moved in the non-conservative direction with the largest difference 2.7 orders of magnitude.

However, for weaker contributors or profiles where there are close proportions and ambiguity, false exclusions or results that favour exclusion can occur. This is not unexpected. Under N-2, fewer genotype combinations result from the deconvolution (usually less genotypes with ambiguity i.e. Q alleles). Therefore, this can result in the exclusion of true contributors present at low levels if their genotype is not proposed and accepted. This also results in more exclusionary *LR*s for known non-contributors, because there are less genotypes with ambiguity in the deconvolution.

In addition, prior to the interpretation of a profile, STRmix™ will attempt to generate a number of genotype combinations with the given profile and assumed number of contributors. If the number of peaks that can only be explained as an allelic peak exceeds two times the assumed number of contributors, STRmix™ will not be able to interpret the profile and will issue a warning. For example, if there are four allelic peaks at a given locus, STRmix™ will be unable to interpret the profile under the assuming of N = 1. The least number of contributors will have to be 2.

Figure 2: Comparison of log(*LR*) obtained when profiles were run in STRmix™ as N versus N+2 contributors for the known and non-contributors (where H_p true is represented by blue circle data points and H_a true are represented by red circles).

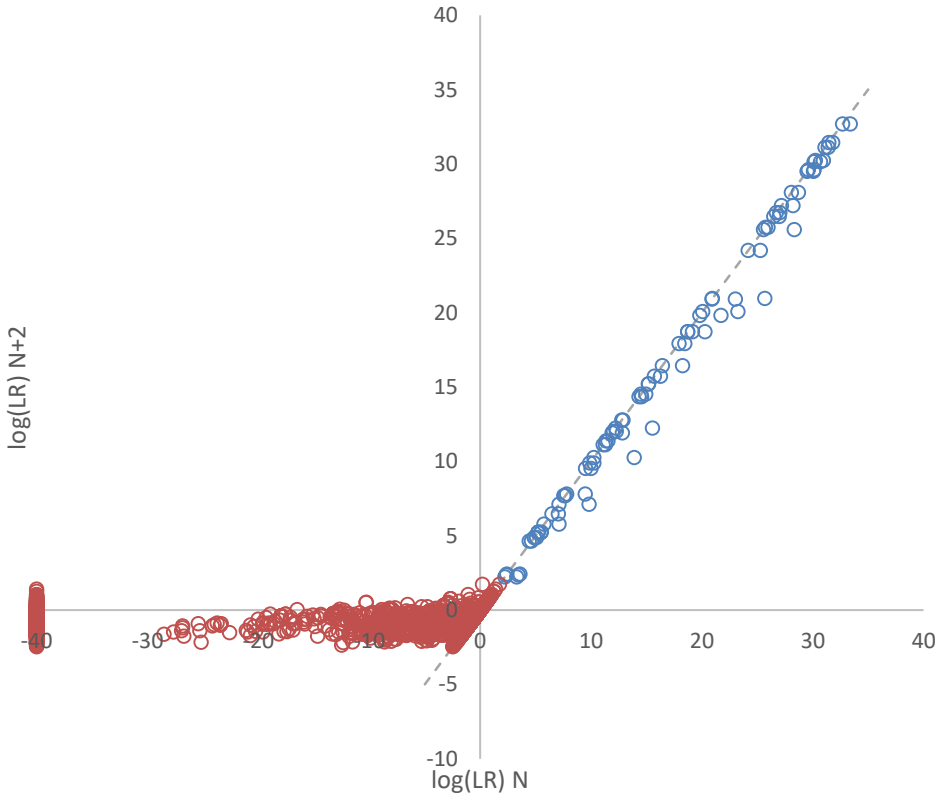

Figure 2 demonstrates that, in general, there is no significant effect on the *LR* of the contributors to the mixture if the number of contributors is overestimated by two additional contributors to the mixture. This is demonstrated by the majority of the *LR*s for known contributors remaining close to the $x = y$ line. However, there are some *LR*s for known contributors which decrease (*LR* approaching 1) when the number of contributors is overestimated. This is because STRmix™ is adding the additional (unseen) profile at low DNA amounts (template), diffusing the genotype probabilities. This also affects the *LR*s assigned to known non-contributors, because more genotype combinations are accepted at a very genotypic weight, resulting in the *LR*s also approaching one. This is shown by the red circles all above the $x = y$ line. These are the expected results.

Some interpretations did not complete under the assumption of N+2 contributors due to limitations of computing power.

References

1. Bright, J.-A., et al., *Searching mixed DNA profiles directly against profile databases.* Forensic Science International: Genetics, 2014. **9**: p. 102-110.
2. Bright, J.-A., J.M. Curran, and J.S. Buckleton, *The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation.* Forensic Science International: Genetics, 2014. **12**: p. 208-214.
3. Moretti, T.R., et al., *Internal validation of STRmix for the interpretation of single source and mixed DNA profiles.* Forensic Science International: Genetics, 2017. **29**: p. 126-144.

4. Bright, J.-A., et al., *Internal validation of STRmix™ – A multi laboratory response to PCAST.* Forensic Science International: Genetics, 2018. **34**: p. 11-24.
5. Bille, T.W., et al., *Interpreting a major component from a mixed DNA profile with an unknown number of minor contributors.* Forensic Science International: Genetics, 2019. **40**: p. 150-159.
6. Alfonse, L.E., et al., *A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt.* Forensic Science International: Genetics, 2018. **32**: p. 62-70.