Forensic Science International: Genetics Supplement Series 7 (2019) 92–94

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

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

How to avoid driving DNA caseworkers crazy: CaseSolver, an expert system to investigate complex crime scenes

Lourdes Prieto[a,*], Peter Gill[b,c], Øyvind Bleka[b]

[a] Instituto de Ciencias Forenses, Universidade de Santiago de Compostela, Spain
[b] Department of Forensic Sciences, Oslo University Hospital, Oslo, Norway
[c] Institute of Clinical Medicine, University of Oslo, Oslo, Norway

ARTICLE INFO

Keywords:
STR DNA mixtures
LR
Expert system
Casework

ABSTRACT

DNA analyses can be used for both investigative (crime scene-focused), or evaluative (suspect-focused) reporting. Investigative, DNA-led exploration of serious crimes always involves the comparison of hundreds of biological samples submitted by the authorities for analysis. Crime stain comparisons include both evidence to evidence profiles and reference to evidence profiles. When many complex DNA results (mixtures, low template LT-DNA samples) are involved in the investigation of a crime, the manual comparison of DNA profiles is very time-consuming and prone to manual errors. In addition, if the person of interest is a minor contributor, the classical approach of performing searches of national DNA databases is problematic because it is realistically restricted to clear major contributors and the occurrence of masking and drop-out means that there will not be a definitive DNA profile to perform the search with.

CaseSolver is an open source expert system that automates analysis of complex cases. It does this by three sequential steps: a) simple allele comparison b) likelihood ratio (LR) based on a qualitative model (forensim) c) LR based on a quantitative model (EuroForMix). The software generates a list of potential match candidates, ranked according to the LRs, which can be exported as a report. The software can also identify contributors from small or large databases (e.g., staff database or 1 mill. individuals). In addition, an informative graphical network plot is generated that easily identifies contributors in common to multiple stains. Here we describe recent improvements made to the software in version v1.5.0, made in response to user requirements during intensive casework usage.

1. Introduction

The usual workflow in DNA forensic casework involves several steps at the laboratory: reception of the samples, analysis, comparisons of profiles, uploads to the national database and LR calculations. Regardless of the phase of the interpretation process (investigative or evaluative), in many labs, part of the workflow is performed in an automated way, but the comparison of profiles is still done manually (either to select which profiles are going to be included in the database or to report the results after the analysis). The possibility of allele drop-out and drop-in complicates the allele match comparison between profiles, mainly for mixtures. One of the possible contributors to a mixture can go unnoticed, resulting in undesirable errors that can adversely affect the investigation. Also, the analyst could report a mixture as inconclusive since it can be challenging to compare it with reference samples. CaseSolver (CS) [1] avoids loss of results that could be useful for the investigation (by decreasing the occurrence of inconclusive results).

In this paper we present an improved version of CS (v1.5.0), an open source expert system that automates analysis of complex cases.

2. Materials and methods

Profiles are imported into CS from a GeneMapper™ table (alleles and peak heights). At this step, CS defines single source profiles and mixtures and compares references to single source evidence profiles. This allows the user to distinguish profiles that are single source from those that are potential mixtures and to discover if unknown profiles are present in the case. A graphical representation of the profiles can be also obtained.

For mixture comparison, CS follows the method presented by Bleka et al. [1] where a step-wise sequential analysis is performed:

* Corresponding author.
E-mail address: lourditasmt@gmail.com (L. Prieto).

https://doi.org/10.1016/j.fsigss.2019.09.036
Received 5 September 2019; Accepted 22 September 2019
Available online 23 September 2019

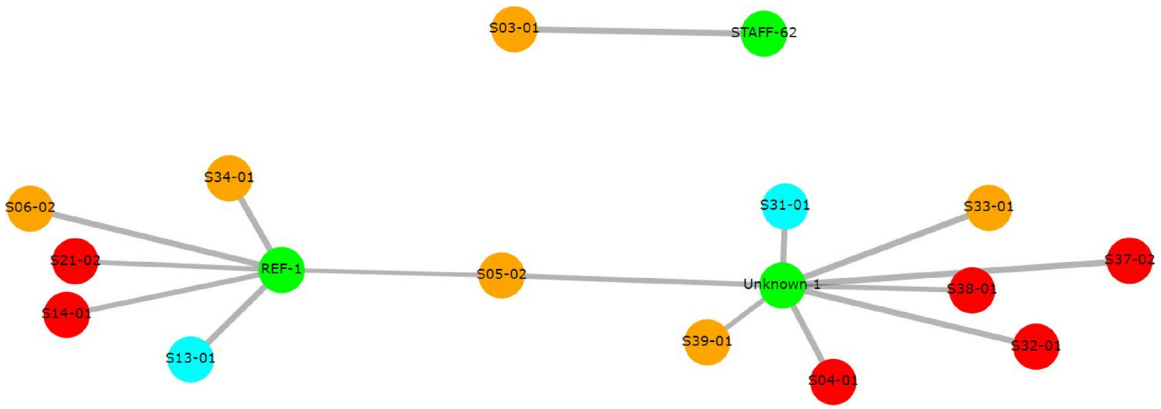

Fig. 1. Graph plot generated from the final match list of a case involving 40 evidence samples (15 DNA mixtures and 25 single source profiles) and 1 reference sample (REF-1), before performing deconvolution. Green nodes are reference profiles, cyan nodes are single source profiles, whereas the orange and red nodes are mixtures; orange nodes are mixtures with 2 contributors, and red nodes are mixtures with more than 2 contributors (estimated). The majority of single source profiles are not shown in the graph for simplicity to the reader. Only 12 mixtures are in the graph since three of them did not match any reference (candidates were obtained after deconvolution). Note that S03-01 is a 2-person mixture matching the profile of one person of the laboratory staff (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

a) First a simple allele comparison is undertaken. The proportions of the alleles of the reference that are included in the mixtures are shown. The user can set the matching allele count (MAC) threshold (e.g. 0.9 corresponds to 10% of mismatch between evidence and reference profiles). Candidates with an allele proportion higher than the MAC threshold go to the next step.
b) Second, an analysis with a qualitative model (forensim) [2] is carried out. The software estimates the number of contributors and calculates the qualitative LRs (Qual.LR), where the propositions are: Hp: "Person of interest (POI) and K-1 unknowns (unrelated) are contributors to evidence profile" and Hd: "K unknowns (unrelated) are contributors to evidence profile E". The user can set a Qual.LR threshold (e.g., LR = 10) to define the candidates that will go to the next step. The time of computing at this step varies depending on the profiles involved but it is usually very short (seconds).
c) Finally, a quantitative model (*EuroForMix*) [3] is used in order to provide the ultimate list of candidate matches. The LR for each candidate passing the threshold defined from the second step (Qual.LR) is re-calculated based on the quantitative model (Quan.LR) with the same propositions as for Qual.LR. The user can define a Quan.LR threshold from which the final list of candidates is shown (e.g., LR = 1000). The time of computing this step varies depending on the profiles involved but it takes usually several minutes. The user may also choose to skip this step and instead manually re-calculate relevant comparisons to reduce the time.

In addition to calculating LR, CS is also capable of performing deconvolution based on the quantitative model and (suboptimal) testing for relatedness by comparing the number of shared alleles between references that are identical by state (IBS). The software also provides a graphical network of the matched samples, as well as a comprehensive report.

Since the previous CS version (1.4.3), several changes and useful features have been added to the new version (1.5.0). Ten cases involving between 50–120 evidence samples (always including DNA mixtures, partial and complete profiles, as well as low level DNA contributors) were tested with this new version.

3. Results and discussion

For a detailed list of the changes performed, the reader can consult the "Changes" paragraph in the CS web page (http://www.euroformix.com/casesolver), but briefly:

a) Some bugs detected in the "Deconvolution" function were fixed.
b) Interactive plots of the graphical representation of the electropherograms were improved using the plotly R-package (also in EuroForMix).
c) The graphs for the Match Status Networks were improved (see Fig. 1). These graphs allow the user to see the matched samples in an interactive graph.
d) The user can choose if all single source profiles are now added to comparison and LR calculations so that the LR can be calculated for each evidence profiles (in the previous version, this was only possible with DNA mixtures).
e) It's now much simpler to perform some function on multiple selected profiles (view data, deconvolve, store, open in EuroForMix)
f) A dedicated SNP module was added (all profiles are assumed to be 3-person mixtures).
g) New saved project files are much smaller now compared to earlier versions.

This new version of CS gives improved profile visualization and a simpler, easier and more flexible profile selection. Also, single source profiles can be compared using LRs (not only mixtures). This is very useful for extremely complex cases involving related individuals. In these cases, a partial single source profile can perfectly (but adventitiously) match a reference, and this would not be discovered if only the simple "comparison of alleles" approach is manually performed. This is also useful if the user wants to run and store the LR on a large set of profiles where some also are single source profiles (typical in validations and research).

4. Conclusions

CS provides an automated overview of all the evidence profiles (both mixtures and non-mixtures), where the peak height information is taken into consideration. The software helps the users to discover potential matches when a large number of samples are analyzed and offers the results in a very simple and graphical way; it deconvolves profiles that can be searched in a database and establishes tentative relationships among profiles. CS is also able to compare profiles of the cases with profiles from the staff. All these features make CS a useful tool to avoid undesirable errors and to answer difficult questions in the complex comparisons (e.g., could an unknown contributor of this 2-person mixture have contributed this other 3-person mixture?).

Declaration of Competing Interest

The authors declare that they have no conflict of interest. There are no direct or indirect financial benefits from any of the open-source software cited.

References

[1] Ø. Bleka, L. Prieto, P. Gill, CaseSolver: an investigative open source expert system based on EuroForMix, Forensic Sci. Int. Genet. 41 (2019) 83–92.
[2] H. Haned, Forensim: an open-source initiative for the evaluation of statistical methods in forensic genetics, Forensic Sci. Int. Genet. 5 (4) (2011) 265–268.
[3] Ø. Bleka, G. Storvik, P. Gill, EuroForMix: an open-source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts, Forensic Sci. Int. Genet. 21 (2016) 35–44.