Forensic Science International: Reports 2 (2020) 100132

Contents lists available at ScienceDirect

Forensic Science International: Reports

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

Forensic Genetics

GENis, an open-source multi-tier forensic DNA information system

Ariel Chernomoretz [a,b,c,*], Manuel Balparda [d], Laura La Grutta [e], Andres Calabrese [e], Gustavo Martinez [f,g], Maria Soledad Escobar [d], Gustavo Sibilla [d]

[a] Dept de Física, Facultad de Ciencias Eaxctas y Naturales, Universidad de Buenos Aires, Argentina
[b] Instituo de Física de Buenos Aires (FIBA) UBA/CONICET, Argentina
[c] Laboratoria de Biología de Sistemas Integrativa, Fundación Instituto Leloir, Argentina
[d] Fundación Sadosky, Argentina
[e] Baufest, Argentina
[f] Servicio de Genética Forense y Registro Provincial de Datos Genéticos, Superior Tribunal de Justicia de la Provincia de Entre Ríos, Argentina
[g] Cátedra de Química Legal, Facultad de Ciencia y Tecnología, Universidad Autónoma de Entre Ríos, Argentina

ARTICLE INFO

Keywords:
DNA database
Autosomal STR
Likelihood ratio
Mixture profiles

ABSTRACT

GENis is an open source multi-tier information system developed to run a forensic DNA database at local, regional and national levels.[1] It was conceived as a highly customizable system, enforcing several security policies including: data encryption, double factor identification, structure of user's roles and permissions, system-wide secure log auditing, non-repudiation protocols and a blockchain-based option to reinforce genetic profiles integrity. GENis is able to perform genetic profile queries of autosomal STR's and its design follows ENFSI[2] and ISFG[3] standards and recommendations. In this work, we present a summary of GENis system design and architecture, the implemented matching rule definitions and the framework used to provide statistical significance to profile matches.

1. Introduction

The unprecedented statistical power of DNA technology as an identification tool has already produced a profound impact in criminal justice. In particular, the creation and development of large DNA databases have unleashed the potential of this technology to solve criminal cases [1,2].

A forensic DNA information system stores different kind of information such as offender's DNA profiles, genetic evidence found at crime scenes and genetic information of victims. In addition, any such system should be able to handle elimination lists of registers pertaining to personnel involved in criminal investigations, DNA-labs or chain of custody tasks. The integration of this information into a computerized environment allows implementing systematic storage and automatized comparisons among DNA profiles. This kind of systems can boost the investigation of crimes by linking DNA profiles from crime-related biological trace material to each other and/or to possible individual contributors [3–5].

There are different IT platforms developed to run national DNA databases world-wide. In a 2016 survey, INTERPOL reported that national DNA databases were already operative in 69 member countries. The software CODIS, developed by the FBI, was reported to be used by more than half of them while other countries adopted their own developed technology to run their facilities [6,7]. In addition, open source solutions like SmartRank have also been released to mine national databases for contributors to complex DNA profiles [8].

GENis is a DNA information system that provides data integration capabilities at regional and/or national level. It is composed by three different modules related to: (a) person identification and analysis of forensic evidence, (b) missing person identification (MPI) and (c) Disaster victim identification (DVI). In this article, we will present a detailed analysis of the functionality implemented in the first module of GENis. We will focus on system design choices, matching rules and statistical models used to assess statistical significance of detected profile matches. The paper is organized as follows: in Section 2 we summarize software design choices for the system's architecture. In Section 3,4 and 5 we present adopted strategies to manage the building blocks of our forensic system: marker kits, allele frequency tables and DNA profiles respectively. In particular, we explain the *group* and *category* classification system implemented in GENis to provide maximum flexibility in connection with stringency and matching rule definitions. In Section 6 we summarize the statistical framework used to weight profile associations. Simulation results were also included in order to quantify the statistical power of the system. Section 7 describes how GENis organizes matching results for active queries and introduce a scenario-testing tool designed to aid the user to weigh different hypothesis. In addition, we summarize in this section the system-wide notification circuit between involved parties to

* Corresponding author.
E-mail address: achernomoretz@leloir.org.ar (A. Chernomoretz).

[1] Fundación Sadosky holds all rights over GENis software and authorizes the publication of this article.
[2] European Network of Forensic Sciences Institutes.
[3] International Society of Forensic Genetics.

http://doi.org/10.1016/j.fsir.2020.100132
Received 18 June 2020; Received in revised form 5 August 2020; Accepted 6 August 2020
Available online 10 August 2020

Table 1
Summary of open source technologies that are used in GENis.

Technology	Context
Slick + PostgreSQL	Profile format, statistics, general queries
MongoDB + Apache Spark	Storage of genetic profiles MapReduce matching algorithms
OpenLDAP	Users, roles, permissions, certificates
AngularJS + Bootstrap	Front end
Akka	Triggering of queries at profile's registering time
Scala + Play!	Reactive architecture

convert a hit into a match. In Section 8 some considerations about GENis multi-tier deployment are presented. Finally, discussion and conclusions are drawn in Section 9.

2. Software architecture

The GENis system exclusively relies on open source technology (see Table 1). The server was implemented in JVM8 and the Scala functional programming language. For security and performance reasons, information is stored in two different databases: a) A relational PostgreSQL database that stores system configuration, DNA profile metadata and the operational log database and b) A non-relational MongoDB database that stores DNA profiles and executes matching queries taking advantage of MapReduce operations.

GENis interface was built following the single-page application pattern. In this type of applications, the browser executes the front-end component following a model–view-controller framework,and communicates with the server through remote services. The web presentation is generated using AngularJS visual components, and Bootstrap for laying them out. The front-end application communicates with the server through REST/JSON services. The server of the application was built on Play! Framework 2.3.x.

We implemented a double authentication strategy to warrant system accessibility: a username and password combination should be provided at login, along with a one-time password generated from a shared secret key and the current time. This time-based one-time password is provided by the Google Authenticator App. Moreover, user's roles can be defined in order to restrain the operations each user can perform in the system. Table SM-T1 in supplementary material lists the kind of permissions that can be granted to different user's roles. Further details on user registration and authentication procedures can be found in Section 2 of the User Manual (UM) included as Supplementary Material.

Additionally, an optional functionality allows checking the integrity of genetic profiles through the generation of cryptographic signatures and their inclusion in an immutable distributed ledger administered, for instance, by a national blockchain authority.[4] That authority can guarantee the immutability of each stored record, in such a way that the authenticity of all active profiles within the database becomes verifiable. Thus, the blockchain optional functionality brings the possibility of verifying that a genetic profile has not been modified since inception in the database.

3. KITS and markers

Most of the STR markers and kits that are currently used in forensic analysis laboratories can be employed in GENis. Currently, 39 autosomal kits are included (see Table SM-T2 for a summary of STR markers and kits). In addition, new STR markers can be included and custom kits can be defined upon them (see Sections UM-8 and UM-9 for further details).

[4] Operational arrangements have already been addressed in Argentina with the national blockchain authority (BFA -Blockchain Federal Argentina) to activate this option upon request of the court system.

4. Allele frequency tables

Statistical significance assessments (e.g. the estimation of random match probabilities) rely on observed allele frequencies reported for the population of interest. GENis provides CRUD capabilities to Create, Read, Update and Delete allele frequency tables that are looked up for probabilistic calculations. Whenever a population allelic frequency table is created or uploaded into the system, users should specify how genotype probabilities should be computed from observed allelic frequencies (the genotype probability model used by GENis is summarized in Section SM-1 of the Supplementary Material). For instance, either the NRC II recommendation 4.1 or the NRC II recommendation 4.10 scheme [9] should be selected, appropriate kinship coefficient values should be specified and a strategy to assign frequency values to rare new alleles should be chosen (see SM-2 for details). Section UM-10 describes implementation and operational details.

5. Profiles

5.1. DNA profile model

GENis uses a binary model to represent DNA electropherograms (see SM-3 for details). DNA profiles can be automatically imported from *GeneMapper's* (Applied Biosystems, USA), combined-table exported text files. Alternatively, profiles can be introduced manually using a double-blind procedure. Metadata, such as court case information, type of biological sample, etc, can also be input into the system. Importantly, every entered profile should be assigned to a given profile-group and profile-category. These features lay down important aspects of the profile usage and management inside GENis, like admissibility, stringency criteria and specific matching heuristic rules (see next Section).

Quality assurance policies for DNA-profiles are necessary to promote a reliable database performance (see Sec 3.5 and 3.9 of [4]). Definitions and recommendations provided by expert committees about the admission criteria for low template DNA samples and/or mixtures of more than 2 or 3 contributors, can be accommodated and implemented in the system. Apart from the chosen criteria, GENIs follows the strategy proposed by Haned and collaborators [10] to estimate the number of contributors of a given mixture that can be used for subsequent analysis (for the sake of completeness we summarized the followed approach in SM-5).

A given GENis instance could incorporate DNA profiles coming from different laboratories. The system provides CRUD capabilities to manage information from each contributing lab. In particular, estimations of *drop-in*, and *drop-out* parameter values, to be used in LR calculations, could be specified for each lab.

5.2. Profile groups and categories

Profile groups and profile categories are defined in GENis in order to implement a customizable profile organization scheme that permits to manage the system behavior in connection with admissibility criteria, association rules, and matching procedures.

The *group* classification is meant to differentiate DNA profiles coming from stains obtained at crime scenes from those profiles obtained from known sources. The former could be indexed, for instance, as belonging to the ***U**ncertain* group, *U*, whereas the second ones as belonging to the ***C**ertain* group, *C*.

Each profile should be further characterized as belonging to one of system-wide categories that can be defined to fine-tune the logical processing rules that will affect it. The categorization scheme is fully customizable and can be adapted to different working protocols and policies. Table 2 below shows a possible categorization system.

Different admissibility requirements can be independently set for each profile category regarding, for instance: the minimal number of informative loci or the maximal number of loci showing trisomy. In

Table 2
Possible categorization scheme for DNA profiles. n is the estimated/assumed number of contributors.

Group	Category	Description
C	S	DNA profile of a suspect
	V	DNA profile of a victim
	D	DNA profile of putative cross-contaminant contributor
U	E1	DNA profile from crime scene stains (n = 1)
	E	DNA profile from crime scene stains (n>1)
	Ek	DNA profile from crime scene stains (n>1), with known contributors

addition, profile association rules can also be defined, for instance, to link *V*-type profiles with *Ek* profiles for situations where a profiled victim is one of the known contributors to a crime scene mixture evidence.

Matching rules can be defined to customize the queries that are automatically triggered whenever a new profile of given category is incorporated into the system. It is possible to select several target categories against which the new profile should be compared at entry/creation time. For each of them, it is possible to specify the kind of matching algorithm that should be used (see below), the minimal number of loci displaying concordant allelic values, the maximal number of non-concordant markers that will be tolerated and the stringency criterion of the search. GENis follows the three-level stringency scheme (i.e. high, moderate, low) suggested by ENFSI in Section 5.4.2 of [11] (see SM-6 for details).

Every profile included in the database is in an *active* state by default. This means that it will participate in every matching procedure affecting its own category. The removal of a given profile from the database is implemented as a logical procedure that irreversibly changes its status to an *inactivated* one. In this new state the profile is simply ignored for future queries and cannot return to its previous state. Eventually a hard deletion protocol could be incorporated in accordance to specific legal frameworks governing expungement and profile removal polices.

Further functionality and operational details regarding profile groups and categories can be found in Section UM-7.

5.3. Profile matching

Fig. 1 depicts different kind of matching rules that could be established between different profile categories. There are 2 kinds of matching strategies to achieve the following different tasks: person identification (I) and establishment of contributor status in a given evidence (C). In addition, the information of the victim associated profile, *V,* for the *Ev* case is used to define obligated alleles in the search heuristics (A).

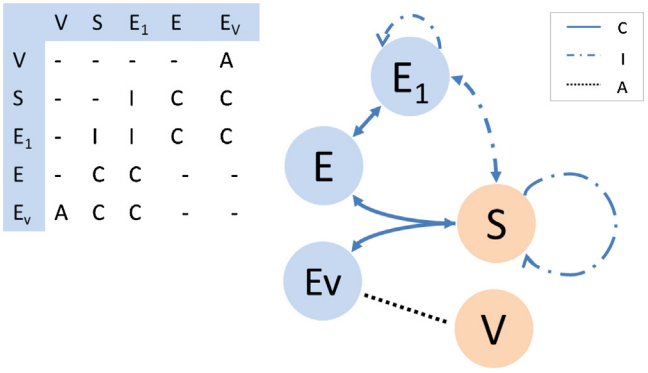

Fig. 1. Purpose underlying match detection between profiles of different categories (described in Table 2). I: identification, C: contribution. A: association link. Right panel: Metagraph of searching/matching rules. Nodes represent profile categories. The dotted line represents an association between *Ev* and *V* categories used to identify *obligated* alleles in search heuristics. Solid arrows represent matching rules with specific stringency criteria to identify contributors of a given evidence (C), and dot-dashed arrows represent matching a strategy to asses for the identification of contributors. The adjacency table of the graph is included as an inset.

When a new profile enters the system, queries are automatically triggered according to the pre-defined rules. For instance, according to the query rules depicted in Fig. 1, a new suspect profile (*S* category) will be compared with already stored profiles of category E_1, *E* and *Ev*. A match will be reported considering stringency criteria (see SM-6) defined for each kind of comparison. In this way, a "high stringency" matching level could be employed for $S<->E_1$ associations, as an identification task lies behind a match detected between profiles of such categories. On the other hand, a "moderate stringency" level would be consistent with the *contribution*-like relationship that exists between S and E category profiles.

6. Genotype probabilities and likelihood calculations

GENis leverages on the statistical framework developed by Curran and collaborators [12] to estimate the probability of observing DNA evidence under different hypothetical scenarios (for completeness purposes, a brief overview of the methodology is included in SM-4). The implementation followed the one considered in LRmix Studio and its R code *forensim* package, a renowned open-source software suite developed for the interpretation of forensic DNA mixtures [13–15].

6.1. Complex DNA profiles

In this section we present a simulation study to assess for the statistical power of GENis methodology to identify contributors to multiple-donor DNA profiles. We considered allele frequencies for 15 autosomal short tandem repeats loci for the American Caucasian population [16]]. CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11 markers belonged to the core CODIS loci used in the US, whereas D2S1338 and D19S433 belonged to the European core loci. We adopted a simulation strategy similar to the one used by Benschop et al. to validate the Smartrank software [8].

Using allelic probabilities, we first simulated, with the aid of the *forensim* R package [13], 25 reference profiles denoted 'seed genotypes'. For each one of them we generated a set of *resembling* genotypes considering almost identical copies of the original profiles, but for a single allele of a randomly chosen locus changed for a *rare allele* (we assumed $P_{rare} = 2.4\,10^{-4}$). We also generated 30 additional profiles that shared 100 % (10), 75 % (10) or 50 % (10) of alleles with two or three-donor mixtures generated using the *seed* genotypes (*from-mix* profiles). We additionally considered, for each *seed* genotype, a parent-like and a brother-like profiles. Finally, 200 independent random genotypes were sampled from the population. Overall, 330 profiles were simulated (25 seed, 25 rare-copies, 50 familial, 30 from mixtures and 200 independent profiles).

As a gold standard we considered ten mixture profiles generated from two and three known seed profiles respectively. Drop-out altered profiles were simulated for each one of them randomly removing 0%, 20 % and 50 % of their allele content to model none, moderate or severe drop-out situations. Each of the 330 profiles of the simulated database were then examined under a prosecutor and defense hypothesis and corresponding LR values were estimated. The true number of contributors, a fixed drop-out rate of 0.01, a drop-in probability of 0.05 and a θ -correction value of 0.01 were considered in the calculations.

Table 3
The table reports the number of sought (S), rare (R), from-mix (M), father-like (F), brother-like (B) and population (P) profiles having LR values greater than unity. Five mixture samples, coming either from two- and three-donors, were considered for a given drop-out level. The AUC_{01} column shows the estimates of area under the ROC curve (see text). Values of this quantity were normalized and lay in the [0,1] interval. A value of unity means that every single sought seed-profile was ranked before the other type of analyzed profiles.

Contributors	Drop-out	S	R	M	F	B	P	$AUC_{0.1}$
2	0%	10 (100 %)	7 (70 %)	1	0	3	0	0.995
	20 %	10 (100 %)	6 (60 %)	1	1	1	0	0.995
	50 %	8 (80 %)	3 (30 %)	0	0	0	0	0.949
3	0%	15 (100 %)	5 (30 %)	0	0	2	0	0.994
	20 %	11 (73 %)	2 (13 %)	0	0	2	0	0.973
	50 %	4 (27 %)	1 (6.7 %)	0	0	0	1	0.922

In Table 3 we reported the number of cases displaying LR values greater than one, as this is a first order minimal necessary condition for profile identification. For two-donor mixtures, a complete retrieval of the right seed-profiles was achieved for none and moderate simulated drop-out levels. On one hand, 20 % of the sought profiles were missed for the severe drop-out situations. The retrieval of seed-profiles for 3-donor mixtures was also successful in absence of drop-outs, albeit some performance degradation was observed for moderate dropout levels. On the other hand, high dropout situations severely impeded the identification process for three-donor mixture. It can also be seen from the table that several *rare* profiles, that differed just by one allele from the queried genotype, also presented LR values larger than unity.

Fig. 2 displays LR estimations for the five 2-donor and the five 3-donor mixtures in the upper-left and upper-right panels respectively. Each subpanel shows data for the three dropout levels considered for each mixture-profile. Red, orange, violet, light-blue, blue and grey colored circles represent seed, rare, from-mix, father-like, brother-like and unrelated non-contributor profiles. Horizontal lines signal the LR = 1 level. Lower left and right panel display ranking values aggregated by drop-out simulated levels for two- and three-donor mixtures respectively. It can be seen for 2-donor mixtures (upper-left panel) that the sought seed genotypes typically presented LR values larger than unity by several orders of magnitude. Some performance degradation can be observed for increasing drop-out levels. However, the right genotypes were typically over-represented in top-ranking LR positions. In order to quantify this trend, we considered results in rank space and estimated the area under the receiver-operator curve (ROC), between 0.9 and 1 specificity boundary levels ($AUC_{0.1}$). The ROC curve serves to illustrate the ability of a binary classifier system (e.g. sought profile or not) in terms of sensitivity and specificity, parameterized by the considered classification threshold (LR value larger than a given value). The normalized partial AUC statistics shown in Table 3 is commonly used as a summary measure of the receiver operating characteristic (ROC) curve. It ranges from 0 to 1, one being a perfect classifier that ranks test cases on top of control cases [17,18]. Despite the sought genotypes still presented the largest LR values for 3-donor samples (bottom-right panel and AUC_{01} values reported in Table 3),

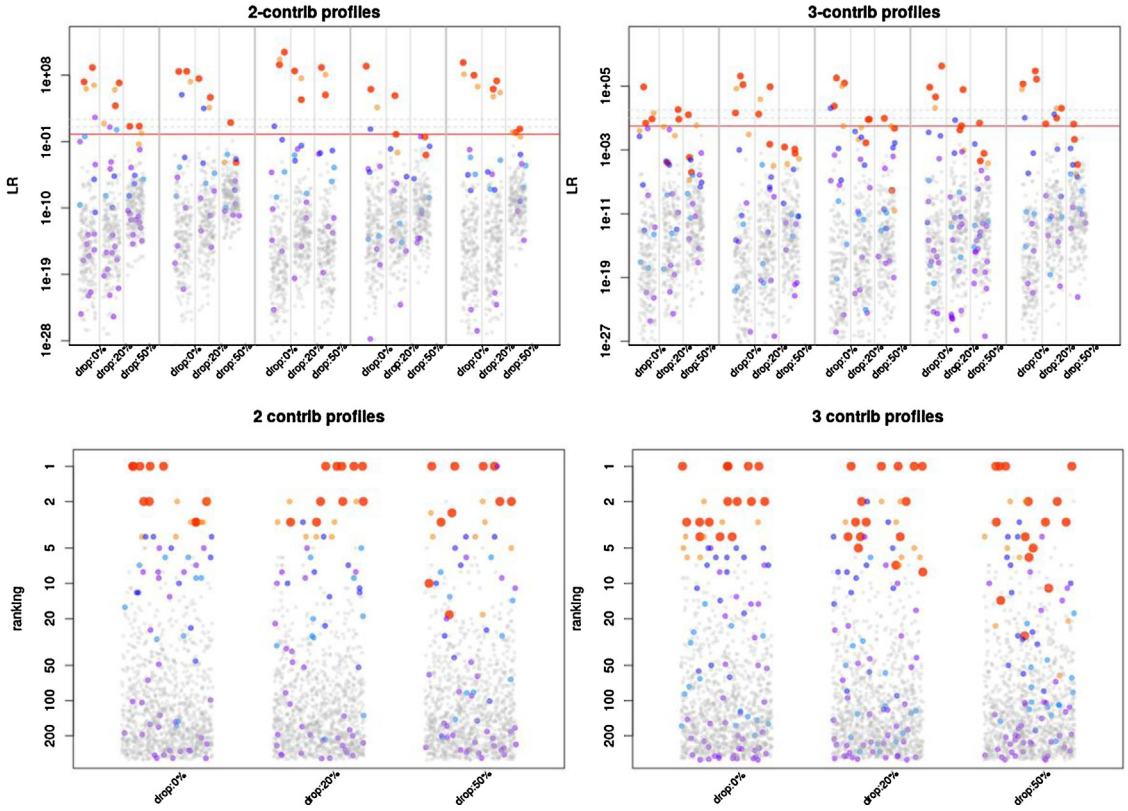

Fig. 2. Upper-left and upper-right panels show LR estimations for the five 2-donors and the five three-donors mixtures respectively. Each subpanel shows data for the three dropout levels considered for each mixture-profile. Horizontal lines signal the LR = 1 level. Lower left and right panel display ranking values aggregated by drop-out simulated levels for two- and three-donor mixtures respectively. Red, orange, violet, light-blue, blue and grey colored circles represent seed, rare, from-mix, father-like, brother-like and population independent profiles respectively.

Table 4
Analysis aim and default statistic employed for each query-target profile category combination (shorthand notation is used for LRs). Q: query profile, T: target profile, V: certain contributor profile (i.e. victim), U_i: i-unknown contributor, n_Q (n_T): number of contributors of the query(target) profile.

Query profile category	Target profile category	Default statistic	Looking for	Scenario testing
S	S	$LR = \frac{1}{U_1}$	identical twins db consolidation	No
S	E_1	$LR = \frac{Q}{U_1}$	identification	No
S / E_1	E	$LR = \frac{Q+U_1+\ldots+U_{n_{T-1}}}{U_1+\ldots+U_{n_T}}$	contribution	Yes
	E_V	$LR = \frac{Q+V+U_1+\ldots+U_{n_{T-2}}}{V+U_1+\ldots+U_{n_{T-1}}}$		
E	S / E_1	$LR = \frac{T+U_1+\ldots+U_{n_{Q-1}}}{U_1+\ldots+U_{n_Q}}$		
E_V		$LR = \frac{T+V+U_1+\ldots+U_{n_{Q-2}}}{V+U_1+\ldots+U_{n_{Q-1}}}$		

a general decline of absolute LR values can be observed for these cases (upper-right panel, and column S of Table 3). These findings warn against the use of poor-quality 3-donor samples for identification purposes.

7. Matches and hits management

7.1. The match manager

Much work was devoted to the way results are presented to GENis users. In order to ease the analysis of query results for a given query profile (Q), GENis groups together target-matching profiles (T) by categories. For each one of these groups, hits can be listed by the mean fraction of shared alleles, the number of shared markers and the LR statistics defined to assess the statistical significance of the match (see Section UM-16 for implementation/operational details). Table 4 summarizes which statistics are considered by default for each kind of comparison.

7.2. The scenario-testing tool

Matching results can be prioritized by any of the above mentioned statistics. Noteworthy, GENis provides a *scenario testing framework,* inspired on the LRmix Studio software [13,14] to further analyze reported matches in much detail. Different hypothesis and scenarios involving query (Q), target matching profiles (T) and *certain-group* profiles (C) matching the evidence sample at *moderate-level*, can be tested using this tool (see Section Manual UM-16.6 for usage examples).

7.3. Converting matches into hits

GENis provides a notification tray to inform each geneticist whether a match triggered by a third-party query involved any of his/her profiles (see Section UM-18 for further details on the notification system). In this way, when a potential hit is identified, the responsible geneticists are notified through the system. At this point the involved parties should contact each other to validate or refute the match. The *match manager module* will display the status of the reported match as: *pending* (the match should still be validated by one of the involved parties), *dismissed* (both parties agree on dismiss the match), *confirmed (*by both parties*)* or *in-conflict* (whenever one party confirmed the match, the other have dismissed it). See Section UM-16.4 for further operational details.

8. Multi-tier design

A GENis server can be locally installed to manage a single forensics lab's DNA database. However, the main purpose of the GENis system is to be deployed in a multi-tier hierarchical architecture in order to share information between local, regional and/or national instances of the system.

The GENis tree-like network is built upon Laboratory and Registry type of nodes that are associated to participant forensics labs, where local profile databases are actually stored. These nodes are the" leaves of the tree". Regional Registries serve to integrate the information from affiliated laboratory nodes, and could typically be defined following judicial, geographical and/or administrative basis. In addition, a single master National Registry node can be deployed in order to coordinate and integrate data over the entire network.

Instance interconnectivity allows different nodes of the GENis ecosystem to communicate with each other through the hierarchical tree. Connectivity between instances uses SSL certificates to encrypt communications or can be configured to run over a VPN.

Uploading profiles information to a superior instance can be specified at profile's entry/loading time. When a profile is sent to an upper instance it is stored in the corresponding category (mapping rules can be established to harmonize profile group and category definitions), and automatic queries are triggered in the higher instance. Whenever a match is detected, the information is transmitted to the involved lower instances in order to validate or discard the hit as explained en Section 7.3. Operational details of instance interconnectivity can be found at Section UM-21.

9. Discussion and conclusions

The use of DNA databases has had a profound impact on the ability to identify suspects linked to crime-scene evidence and to suggest/support investigation leads to relate evidence traces of unsolved cases.

In Argentina, following a request of the judiciary, the National Ministry of Science, Technology and Innovation undertook in 2014 an initiative to develop an open source software to assist in criminal investigations by identifying persons through biological evidence. Argentina is a federal country composed of 24 autonomous provinces that enact their respective laws. In the last 15 years, 19 provinces created their own genetic databases with different criteria regarding the types of crimes applicable and the defendant procedural status. With this in mind, GENis had a two-fold objective. First, it was aimed to provide a state-of-the-art tool for judicial institutions for storing and comparing DNA profiles in criminal cases. Secondly, GENis was intended to elicit comments and encourage debate in the experts' community on the implementation of uniform protocols. Therefore, GENis could help to harmonize policies among provinces and facilitate data sharing strategies at the regional, national and international level.

GENis is a DNA information system that provides data integration capabilities at regional and/or national level expanding the scope, significance and capabilities of this kind of systems. the system-wide coordination of forensic information not only provides the possibility of running queries between otherwise independent local instances but also promotes the standardization of data structures and protocols.

As we have shown throughout the article, GENis is a highly flexible information system aimed to implement a very comprehensive DNA-related identification task. From a practical point of view, GENis integrates many ideas already developed by the international community, along with new statistical figures, into a single, unified and highly customizable framework that can accommodate many analysis workflows. In particular, GENis implements a novel methodology, specifically developed for binary profile models, to assess for the statistical significance of the identification of common contributors in DNA mixtures. A full mathematical derivation and a thorough characterization of this novel likelihood-ratio statistic exceed the scope of the present manuscript and will be presented as a separated contribution.

Importantly, the system was entirely developed with strict adherence to open source policies to encourage the application of auditing procedures, and to warrant the integrity and ownership of the stored data. In addition, technical software specifications have been under

exhaustive consistency examination by the Tools and Foundations for Software Engineering Lab [20], of the Computer Department of the University of Buenos Aires.

To date, GENis is already deployed in 18 different Argentinean provinces laboratories, and license agreements have been signed to expand it to new ones. A great effort was made in the training and assistance of technicians, geneticist and IT support teams. We have also made available a website (http://wiki.genis.sadosky.net) from where users can download useful information (handbooks, technical resources) and participate in forums to share operational feedback and suggest further developments.

Declaration of Competing Interest

The authors report no declarations of interest.

CRediT authorship contribution statement

Ariel Chernomoretz: Conceptualization, Methodology, Formal analysis, Software, Writing - original draft, Writing - review & editing. **Manuel Balparda:** Writing - review & editing. **Laura La Grutta:** Software, Formal analysis. **Andres Calabrese:** Software, Resources. **Gustavo Martinez:** Conceptualization, Validation. **Maria Soledad Escobar:** Conceptualization, Project administration. **Gustavo Sibilla:** Conceptualization, Supervision, Funding acquisition, Writing - review & editing.

Acknowledgments

The initial requirement for GENis functionality map was set by the Argentine Society of Forensic Genetics (SAGF), the National Council of General Attorneys and the Federal Council of Criminal Policy. The program was developed by Fundacion Sadosky [21], a public-private institution in the orbit of the Argentine National Science and Technology Ministry (MINCyT), which coordinated an interdisciplinary team integrated by IT professionals, physicists, lawyers, molecular biologists and geneticists, representing very diverse institutional belongings. One remarkable asset of the project has been the synergy achieved through an active public-private cooperation along the whole development process. Researchers from various national universities (Buenos Aires, La Plata, Rosario, Córdoba, Tucumán, Quilmes, Comahue) have participated in different tasks, from the design of statistical models to the testing and validation of draft versions. Other R&D centers, like CONICET, Instituto Leloir Foundation and the Argentine Bioinformatics Society (A2B2C) have also taken part in the project bringing very useful contributions including training and technical guidance. For it's part, the private sector was represented by a local IT company (Baufest) that was in charge of system design and coding. Additionally, the project benefitted from a continued support from the main national IT chambers (CESSI and CICOMRA).

Several Judicial institutions have also cooperated in the system development, including the Federal Court Board (JUFEJUS), the Laboratory of DNA Comparative Analysis of the Supreme Court of Justice of the Buenos Aires Province, the Forensic Genetic Service of the Supreme Court of Entre Rios Province, the Criminalistics and Forensic Sciences Research Institute within the scope the General Attorney's Office of the Province of Buenos Aires, the Genetic Digital Fingerprint Service of the University of Buenos Aires and many others.

We appreciate the financial support from the MINCyT and Fundacion Sadosky. We would like to thank Mario Adaro (Institute of Technology of the Federal Court Board), Andrea Colussi (Forensic Medical Office of the National Supreme Court of Justice), Mercedes Lojo (DNA Comparative Analysis Laboratory of the Supreme Court of Justice of the Buenos Aires Province), Mariana Herrera, Walter Bozzo and Franco Marsico (National Bank of Genetic Data), Daniel Corach and Andrea Sala (Genetic Digital Fingerprint Service of the University of Buenos Aires), Cesar Guida and Elina Francisco (Criminal Investigation and Forensic Sciences Institute of the General Attorney's Office of Buenos Aires province), Cecilia Miozzo (NOA Regional Forensic Genetic Laboratory of the Supreme Court of Justice of Jujuy province), Alejandra Guinudinik (Forensic Molecular Biology Service of the General Attorney's Office of Salta province), Pedro Villagran (Forensic Investigations Center of the General Attorney's Office of Salta province), Silvia Vanelli Rey (North Patagonia Regional Forensic Genetic Laboratory of the General Attorney's Office of Rio Negro province), Cecilia Bobillo (Forensic Genetic Laboratory of the General Attorney's Office of La Pampa province), María Beatriz Vazquez (Forensic Investigations Laboratory of the Supreme Court of San Juan province), Haydee Fariña (Forensic Medical Service of the Supreme Court of Neuquen province), Gabriela Lamparelli (Forensic Sciences Institute of the Supreme Court of Chaco province), Juan José Belzki (Forensic Investigation Center of the Supreme Court of Formosa province), Agustina Dorigón Lezana (Forensic Genetic Cabinet of the Supreme Court of Santiago del Estero province), Victor Moloeznik (General Attorney's Office of Santa Fe province), María Consuelo Martí (Chemical Laboratory of the General Attorney's Office of Santa Fe province), Juan José Galvez (Forensic Medical Institute of the Supreme Court of Corrientes province), Diego Rinaldi (General Attorney's Office of Corrientes province), Noelia Massari (Regional Forensic Investigations Laboratory of the General Attorney's Office of Chubut province), Miguel Rubio (Forensic Genetic Cabinet of the Supreme Court of Tucuman province), Nestor Basso and Lidia Vidal Rioja (CONICET), Hortensia Cano (South Patagonia Regional Forensic Genetic Laboratory of the Supreme Court of Justice of Santa Cruz province), Inés Aparici and Jessica Name (Supreme Court of Tierra del Fuego Province), Cristina Buslje (Fundación Instituto Leloir, A2B2C), Sebastian Uchitel, Hernán Melgratti and Víctor Braberman (Buenos Aires University, CONICET), Gustavo Parisi (National Quilmes University), Paula Venosa (La Plata University), Sandra Furfuro (Cuyo University), Manuel Aybar (Tucumán University), Carlos de la Vega (Cordoba University), Juan Luzuriaga and Agustina Buccella (Comahue University) and many others for fruitful discussions.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsir.2020.100132.

References

[1] P.D. Martin, H. Schmitter, P.M. Schneider, A brief history of the formation of DNA databases in forensic science within Europe, Forensic Sci. Int. 119 (2001) 225–231.
[2] J. Butler, Chapter 2—DNA Extraction Methods, Advanced Topics in Forensic DNA Typing, Elsevier, Maryland, 2012, pp. 29–47.
[3] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, Academic press, 2011.
[4] E.Dw. group, DNA Database Management Review and Recommendations, ENFSI, 2016.
[5] P. Gill, Misleading DNA evidence: reasons for miscarriages of justice, Int. Comment. Evid. 10 (2012) 55–71.
[6] G.S. Morrison, F.H. Sahito, G. Jardine, D. Djokic, S. Clavet, S. Berghs, C.G. Dorny, INTERPOL survey of the use of speaker identification by law enforcement agencies, Forensic Sci. Int. 263 (2016) 92–100.
[7] J.M. Butler, S. Willis, Interpol Review of Forensic Biology and Forensic DNA Typing 2016-2019, Forensic Science International: Synergy, 2020.
[8] C.C. Benschop, L. van de Merwe, J. de Jong, V. Vanvooren, M. Kempenaers, C.K. van der Beek, F. Barni, E.L. Reyes, L. Moulin, L. Pene, Validation of SmartRank: a likelihood ratio software for searching national DNA databases with complex DNA profiles, Forensic Sci. Int. Genet. 29 (2017) 145–153.
[9] N.R. Council, The Evaluation of Forensic DNA Evidence, National Academies Press, 1996.
[10] H. Haned, L. Pene, J.R. Lobry, A.B. Dufour, D. Pontier, Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? J. Forensic Sci. 56 (2011) 23–28.
[11] D. ENFSI, DNA Database Management, (2016) .
[12] J. Curran, P. Gill, M. Bill, Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure, Forensic Sci. Int. 148 (2005) 47–53.

[13] H. Haned, Forensim: an open-source initiative for the evaluation of statistical methods in forensic genetics, Forensic Sci. Int. Genet. 5 (2011) 265–268.
[14] H. Haned, K. Slooten, P. Gill, Exploratory data analysis for the interpretation of low template DNA mixtures, Forensic Sci. Int. Genet. 6 (2012) 762–774.
[15] P. Gill, H. Haned, A new methodological framework to interpret complex DNA profiles using likelihood ratios, Forensic Sci. Int. Genet. 7 (2013) 251–263.
[16] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, Allele frequencies for 15 autosomal STR loci on US Caucasian, African American, and Hispanic populations, J. Forensic Sci. 48 (2003) 908–911.
[17] D.K. McClish, Analyzing a portion of the ROC curve, Med. Decis. Mak. 9 (1989) 190–195.
[18] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (2006) 861–874.
[20] D.M. Powers, Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation, (2011) .
[21] The Laboratory on Foundations and Tools for Software Engineering. https://lafhis.dc.uba.ar/home.