International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:11, No:3, 2017

Development of a Software System for Management and Genetic Analysis of Biological Samples for Forensic Laboratories

Mariana Lima, Rodrigo Silva, Victor Stange, Teodiano Bastos

***Abstract*—**Due to the high reliability reached by DNA tests, since the 1980s this kind of test has allowed the identification of a growing number of criminal cases, including old cases that were unsolved, now having a chance to be solved with this technology. Currently, the use of genetic profiling databases is a typical method to increase the scope of genetic comparison. Forensic laboratories must process, analyze, and generate genetic profiles of a growing number of samples, which require time and great storage capacity. Therefore, it is essential to develop methodologies capable to organize and minimize the spent time for both biological sample processing and analysis of genetic profiles, using software tools. Thus, the present work aims the development of a software system solution for laboratories of forensics genetics, which allows sample, criminal case and local database management, minimizing the time spent in the workflow and helps to compare genetic profiles. For the development of this software system, all data related to the storage and processing of samples, workflows and requirements that incorporate the system have been considered. The system uses the following software languages: HTML, CSS, and JavaScript in Web technology, with NodeJS platform as server, which has great efficiency in the input and output of data. In addition, the data are stored in a relational database (MySQL), which is free, allowing a better acceptance for users. The software system here developed allows more agility to the workflow and analysis of samples, contributing to the rapid insertion of the genetic profiles in the national database and to increase resolution of crimes. The next step of this research is its validation, in order to operate in accordance with current Brazilian national legislation

***Keywords*—**Database, forensic genetics, genetic analysis, sample management, software solution.

I. Introduction

A. Forensics Genetics

SINCE the introduction of DNA profiling, in the late 1980s, the advent of DNA fingerprinting revolutionized the forensic science and criminal investigation. Currently, Forensic DNA is considered the "gold standard method" of forensic science, which allows effective human identification due to its high sensitivity [1], [2].

The technology of polymerase chain reaction (PCR) has made possible to amplify small amounts of DNA represented at short tandem repeats (STR). STRs are DNA segments composed of varying numbers of core repeats that are used as genetic markers and become the mainstay of current forensic identity profiling [1], [2].

Mariana Lima, Victor Stange and Teodiano Bastos are with the Postgraduate Program in Biotechnology, Federal University of Espirito Santo, Vitoria, Brazil (e-mail: lugon.mariana@gmail.com, vitustange@gmail.com, teodiano.bastos@ufes.br).

Rodrigo Silva is a system designer from Vitoria, Brazil (e-mail: rodrigooriani@yahoo.com.br).

B. DNA Samples, STR Profile Interpretation and Analysis

The growing number of forensic DNA samples, from both criminal investigations and convicted offender database, produce large amounts of STR profile data, which to be stored and analyzed, requiring significant resources, in terms of equipment and personnel [3]. This requires significant resources, in terms of equipment and personnel. Some advances have been made on development of robotic equipment to automate the extraction, quantification, amplification of DNA and capillary electrophoresis (CE) instrumentation. Other area that currently has received more emphasis is the data analysis stage. Part of this time-consuming process, in recent years, has been the replacement of human workers for bioinformatics software called "expert systems". This kind of software has been developed to identify peaks and assign alleles from profiles originated by CE without human intervention, and automate this part of analysis (as much as possible) [4]-[6]. A chain of complex decisions rules is required to differentiate between biological and technological artefacts from noise in the data. Once the artefacts have been removed, all remaining peaks are assumed to be "real" and the STR profile is established [4].

After the STR profile analysis, forensic interpretation involves the comparison of the DNA profile obtained from another sample or a database to determine if there is a genetic match [4]. DNA database enables the comparison of a huge amount of profiles to try to identify the perpetrator [2].

Other capability of DNA is its inheritance pattern, where half of an individual's genetic load comes from the mother and the other half comes from the father. This allows parents to be used as reference points. Missing persons and other not identified victims can be identified by analysis of STR profiles trough the establishment of kinship associations [2].

C. LDNACr

LDNACr is a software solution for forensic genetics laboratories that are faced to the processing and analysis of an increasing number of samples. It provides toolboxes for automatic management of samples, criminal cases and local databases, allowing sample tracking and minimizing of time consuming in the workflow.

Other important function of this software is the interpretation and analysis of the profiles. After genetic profiles are generated and inputted on LDNACr, the STR markers may be compared and applied to statistical methods commonly used in forensics genetics to establish a statistical interpretation of the results [7].

The next sections describe the details of the software development.

II. Methods

A. Development of the Software LDNACr

LDNACr is a web application, in which the server part runs on a remote computer (connected to the internet), and the client part runs on web browsers from a computer, in order to render and visualize the interface of the application [8]. The user interface (UI) of LDNACr is user-friendly and is written in HTML and CSS languages. CSS (style language) is a commercial web template, named "Inspinia" [9].

Server side code is usually responsible for retrieving and storing data, and, in this case, the language JavaScript with NodeJs platform is used. NodeJS is a server-side platform that is capable of very quickly transferring data, and with efficient storage and performance [10].

B. The UI of LDNACr

The first step on development of LDNACr involves requirements gathering. All the work-flow in a forensic genetic laboratory was based on a forensic genetics laboratory inside the Civil Police of Espirito Santo State (Brazil). In Brazil, most of forensic genetic laboratories do not use software for statistical analysis or sample and local database management, as they use only a national database (CODIS). The functionalities for a software system are: Authentication (security), sample database manager, cases manager (analysis and results), and manager of local profile databases. After functionalities are defined, the development of the UI was initiated.

The software initiates with security screen for user authentication. In this software, the web environment needs to be protected and accessible only by legally authorized people. The data have to be stored in a secure information resource, following recommendations of the National Academy of Science's National Research Council (NRC) in a Committee on DNA Technology in Forensic Science [7].

Once the user makes the login, the home screen offers four sections for access: "Biological Samples", "Exams", "Statistical", and "Local Profiles Database".

The first section is "Biological Samples", where the user can insert a new sample or consult a sample from the database. This section includes forms where the user can input all data related to the sample, which will remain stored in the database. In this same form, there is a label with the information about the sample location on the physical storage. In this stage are made procedures guaranteeing the identity of the origin (individuals/suspect) and data from where the sample is taken (crime/victim). LDNACr offers a tool for data organization and storage, allowing posterior traceability [7]. Fig. 1 shows a form for register a sample that enters in the laboratory.

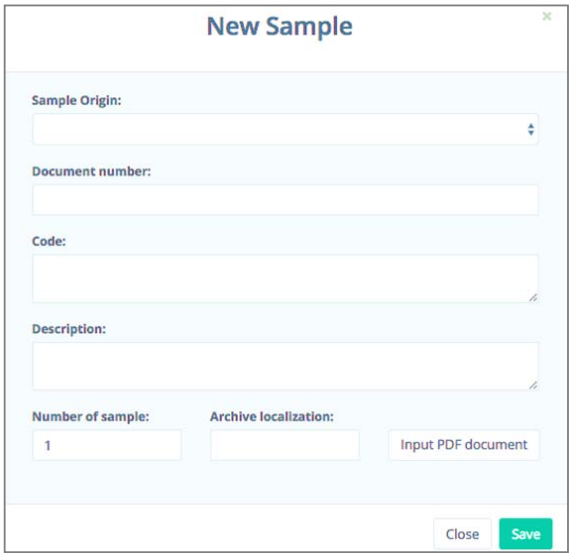

Fig. 1 Screenshot of a "New Sample" registration form

The second section is "Exams" where the user can input all data related to the processing and analysis of samples. In this part the user can detail data from the DNA extraction, PCR amplification and CE. In this same section, it is possible to manage and store data from forensic cases.

This section has a subdivision where the STR profile, obtained after CE and electropherogram analysis, could be inputted. The feature "Analysis and Results" makes the comparison and the statistical interpretation of selected STR profiles.

C. "Analysis and Results"

This feature makes the Likelihood Ratio (LR) comparison of profiles and the statistical interpretation used in forensics genetics, defined by the NRC [7].

The first step is to input the STR profiles from a case, obtained after the electropherogram analysis in .csv file format. It is possible to make the input of all profiles of the same forensic case, which is stored. In LDNACr, a STR profile is represented by the name of the sample, and all alleles from STR regions are amplified through the PCR amplification kit.

Fig. 2 shows part of an STR profile such is displayed in LDNACr.

Sample	AMEL	CSF1PO	D10S1248	D12S391	D13S317
Known Sample	X, Y	10, 13	14, 17	18, 19	8, 9

Fig. 2 Example of part of an STR profile from a known sample

In forensic cases the comparison is basically made between two samples; in most of cases with a suspect (known sample,

K) and the crime scene evidence (question sample, Q), or between two crime scene evidences. The process of comparing STR profiles is limited to one of the tree possibilities:

- Inclusion (Match): STR profiles have the same genotypes, and no unexplained differences exist between the samples.
- Exclusion (Non-match): Profile shows differences that can only be explained by the two samples originated from different sources.
- Inconclusive: Insufficient information exists to support any conclusion, such as poor quality evidentiary samples or lack of a reference sample for comparison purposes.

Once a STR match is done between profiles from a suspect (known sample, K), and from crime scene evidences (question sample, Q), it is necessary to quantify the evidentiary value of this match, which is done using (1) [7].

$$Likelihood\ Ratio = \frac{Hyphothesis\ of\ prosecution\ (Hp)}{Hyphotesis\ of\ defense\ (Hd)} \quad (1)$$

In LDNACr, the comparison between two STR profiles is made when the user selects them. The two profiles are exposed, and, if they match, therefore LR is calculated.

The LR formula used in forensic genetics is based on a comparison of the probabilities of the evidence under two alternative propositions, mutually exclusive, that represents the position of the prosecution (the DNA from the crime scene originated from the suspect) and the position of the defense (the DNA coincidently matches the DNA of the defendant and is from an unknown person out in the population at large). The first hypothesis (the hypothesis of prosecution) is placed in the numerator of LR (1), while the second hypothesis (that someone other than the defendant committed the crime - the hypothesis of defense) is placed in the denominator. A random person is someone in the same population who is not related to the suspect [7].

When there are two STR profiles that match, the hypothesis of the prosecution is that the defendant committed the crime, with 100% of probability, so the denominator is equal to 1 (H_p =1). On the other hand, the hypothesis of the defense that the profile originated from someone else can be obtained from genotype frequencies of a specific STR profile for different populations [7]. These values must be inserted on LDNACr, which uses them to estimate DNA profile frequencies, using the product rule, and calculates LR in the case of match between two samples. For example, in a match, considering one STR region TH01, the genotype was 6,6 with a frequency value 0,139 for 6 alelle, which agrees with the Brazilian population data frequency [11], the LR considering just this one marker is exemplified (2). The calculation is made for all STR regions.

$$LR = \frac{1}{p^2} = \frac{1}{(0{,}212)^2} = \frac{1}{0{,}044} = 22{,}727 \quad (2)$$

LR numbers can be rather large, when the calculation involves all STR analyzed. This is often preferable to report them in terms of their logarithm (log) value [7]. This feature allows rapid LR calculation and the weight of evidence can be effectively evaluated as shown in Fig. 3.

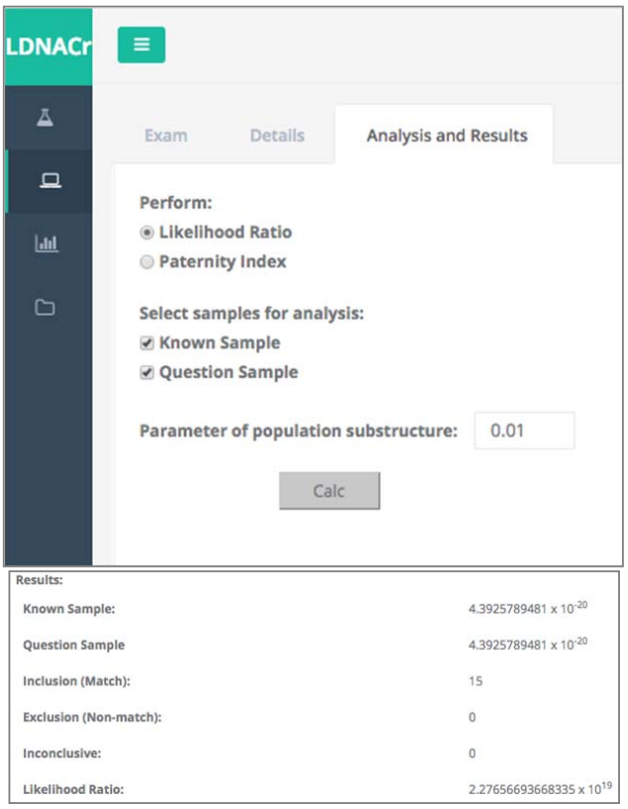

Fig. 3 Screenshots of section "Analysis and Results" showing a LR calculation and results

III. Results and Discussion

This work introduced the LDNACr system, an in-house solution, user-friendly software that handles all the information digitally with high security and minimum manual paper work.

LDNACr aims to provide a powerful tool to reduce time-consuming in workflow and a great database to store all data related to samples that enter in a forensic genetics laboratory, allowing posterior traceability.

Considering statistical calculation, the system provides other statistical feature to calculate the paternity index in addition to the LR calculation, allowing to solve cases involving human identification, such as unidentified victims, mass disaster victims and paternity that involves sex abuse.

LDNACr also provides a local STR profile database and saves all profiles generated by the laboratory. A validation of LDNACr will be performed before use.

IV. Conclusions

Literature shows that in recent years, the advances in automation and software in forensics genetics will increase. There are several software systems available with different functions, but LDNACr is available in national language

(Portuguese), is easy to use, and stores its data in a relational database. LDNACr has as purpose to include almost all the necessity of a forensic genetic laboratory in all-in-one system, such as: sample database manager, cases manager (analysis and results), local profiles database, and statistical features. LDNACr aims to be the first Brazilian software in this area.

Acknowledgment

The authors would like to thank the Postgraduate Program in Biotechnology (PRPPG/Brazil) and the Federal University of Espirito Santo (UFES/Brazil) for their support for this research.

References

[1] D. G. B. Leonard (Ed). "Molecular pathology in clinical practice". Second Edition. *Springer Science+ Business Media*, pp 793-810, 2016.
[2] J. M. Butler. "The future of forensic DNA analysis." Phil. *Trans. R. Soc. B.* vol: 370, pp. 1-10, 2015.
[3] M. M. Holland and W. Parson. "GeneMarker® HID: A reliable software tool for the analysis of forensic STR data." *Journal of forensic sciences*. vol. 56, pp 29-35, 2011.
[4] M. R. Aniba and J. D. Thompson. "Knowledge based expert systems in bioinformatics". *INTECH Open Access Publisher*, 2010. Available from: http://www.intechopen.com/books/expert-systems/knowledge-based-expert-systems-in-bioinformatics Accessed on 14/02/2017.
[5] T. Power, B. McCabe, and S. A. Harbison. "FaSTR DNA: a new expert system for forensic DNA analysis." *Forensic Science International: Genetics*. vol. 2, pp159-165, 2008.
[6] National DNA Index System (NDIS) Operational Procedures Manual, 2017. https://www.fbi.gov/file-repository/ndis-procedures-manual-ver4-approved-04272016.pdf/view Accessed on 14/02/2017.
[7] J. M. Butler. "Advanced topics in forensic DNA typing: interpretation". Academic Press, 2014.
[8] Y. R. Smeets. "Improving the Adoption of Dynamic Web Security Vulnerability Scanners." Master Thesis. November, 2015.
[9] Inspinia. Available from: https://wrapbootstrap.com/theme/inspinia-responsive-admin-theme-WB0R5L90S Accessed on 14/02/2017.
[10] S. L. Bangare, S. Gupta, M. Dalal, A. Inamdarl. "Using Node.Js to Build High Speed and Scalable Backend Database Server." *National Conference "NCPCI-2016"*. pp 61-64, 2016.
[11] V. R. Aguiar, A. M. de Castro, V.C. Almeida, F. S. Malta, A.C. Ferreira, I.D. Louro, I. D. "New CODIS core loci allele frequencies for 96,400 Brazilian individuals". *Forensic Science International: Genetics*, vol.13, pp. 6-12, 2014.