A Tool for Analyzing Mutations in Biomedical Microorganisms

Hector FLOREZ
Universidad Distrital Francisco José de Caldas
Bogotá, Colombia, haflorezf@udistrital.edu.co

Karina SALVATIERRA
Universidad Nacional de Misiones
Posadas, Argentina, kariales@gmail.com

ABSTRACT

Mutations of resistance of drugs are considered a major public health problem. Various mutations associated with the development of resistance of drugs have been described in microorganisms such as virus and bacteria. Analyzing these mutations becomes a challenge because microorganisms are described by a big amount of data that must be processed. Nowadays, scientists in health areas make these analyses manually or using standalone software that provide results in plain text formats, which limit their interpretations. In this paper, we present an online information system for detecting mutations implicated in resistance to drugs. We have used computer applications, technological tools, standard languages, infrastructure systems and algorithms, to analyze positions associated with resistance to different drugs. Said information system allows users to calculate changes in nucleotide and amino acid sequences for each selected sequence from conventional Sanger and cloning sequencing using a friendly graphical interface. In this paper, we present BMA (Biomedical Mutation Analysis), which is the software that we have built to provide sequences analyses with respect to drug resistance. It is available at: http://bma.itiud.org

Keywords: Bioinformatics, mutation analysis, mutation resistance

1. INTRODUCTION

In biology, a mutation is a change in the nucleotide sequence of the genome of an organism, virus, or other genetic elements [1]. Mutations may or may not produce discernible changes in the observable characteristics (phenotype) of an organism. Mutations play an important role in both normal and abnormal biological processes such as: evolution, cancer, and the development of the immune system, including junction diversity [2,3,4].

Mutations can result in several different types of sequence changes; such changes could have no effect, alter the product of a gene, or prevent it from functioning properly or completely [5]. Studying microorganisms that have acquired changes or deletions in their nucleotide sequences is a time-honored practice in biology. Due to mutations can interrupt cellular processes; they often hold the key to understand gene functionalities.

Evolutionary biology suggests that drug selection pressure is an important factor in the emergence and spread of drug resistance [6]. Resistance mutations to drugs are results from amino acid substitutions that produce conformational changes that interfere with drug-target interaction [7]. By altering just one amino acid, the entire peptide may change; thus, the entire protein changes as well. Resistance may involve reduction entry of the drug, changes in the receptor (target) of the drug, or metabolic inactivation of the drug. It is therefore important to determine mutations of resistance to drugs for avoid therapeutic failures. Such a determination is considered an important task for the control and prevent of the emergence of resistance to drugs [8].

The problem about control of infectious diseases is seriously threatened by the steady increase in the number of microorganisms that are resistant to drugs. Resistant infections lead to increased morbidity [6]. The number of treatment failures and drug resistant variants is expected to increase within the next years through selection pressure imposed by drugs therapy.

Several studies have shown that drug-resistant variants that are not detected are often responsible for the failure of a new treatment regimen (e.g., HIV or HCV infections). Monitoring drug resistance is very important for public health. Bacteria have proved adept at developing resistance to new antimicrobial agents, as well as viruses, parasites and fungi.

Sequencing techniques are being increasingly used in the clinical practice to detect drug resistant. Conventional direct sequencing is the gold standard in resistance testing and is used to detect drug-resistance mutations in the molecular targets of therapy, i.e., reverse transcriptase, protease, integrase, and others genes. The sequences obtained by conventional sequencing need to be analyzed quickly and effectively to information that is useful, as is the case with mutations of drug resistance.

Thus, there is an increasing need to develop bioinformatics tools to analyze the rapidly growing amount of nucleotide and amino acid sequence data in different organisms obtained by conventional sequencing [9]. An important task in bioinformatics is the provisioning of data and tools in a simple manner for users to locate and use. Advances in information technology have stimulated the development of new computer applications and algorithms for data analysis, and computer visualization tools for the representation of variation patterns.

The aim of this study was to develop an online information system named Biomedical Mutation Analysis (BMA), which allows users to calculate changes in nucleotide and amino acid sequences for each selected sequence through a friendly graphical interface.

The basic mechanisms in which a microorganism can resist a drug are: a) to alter the receptor for the drug (the molecule on which it exerts its effect); b) to decrease the amount of drug that reaches the receptor by altering entry or increasing removal of the drug; c) to destroy or inactivate the drug; and d) to develop resistant metabolic pathways. Microorganisms can possess one or all of these mechanisms simultaneously. The drug resistance analyses were reviewed carefully. It is based on a comprehensive set of data that were collected from *in vitro* and *in vivo* studies.

2. TOOL FOR ANALYZING MUTATIONS

To create the online information system BMA, we used different standard tools, languages, and infrastructure systems. BMA was designed using the Unified Modeling Language (UML) [10], which allows describing the system following the Object Oriented Paradigm.

Regarding the development of BMA, we used PHP language version 5.3.29[1], which is supported by Apache software version 2.4.7[2] as the application server. For the front end of BMA, we used Bootstrap version 3.3.6[3], which is the most popular HTML, CSS, and JavaScript framework for developing responsive web projects. BMA also has some features based on JavaScript language supported by JQuery version 1.12.3[4], which is a JavaScript library that facilitates some specific JavaScript functionalities. BMA provides three different outputs, where two of them use additional support. The former result is a report generated as a pdf file, which is built using ezpdf version 0.0.9[5], which is a library that supports the creation of pdf files. The latter result is a force-directed graph, which is created using D3 (Data-Driven Documents) version 3.5.16[6], which is an online JavaScript library that helps to deploy data using fancy visualizations.

BMA stores all information related to the mutation analyses in one database supported by MySql version 5.7.12[7], which is a relational database management system. The database includes the entities and relationships required for handling all information related to the proposed mutation analyses. The database is manipulated through project phpMyAdmin version 4.3.11[8], which is software written in PHP intended to handle the administration of data stored in MySql databases. The database was design using the tool MySql Workbench version 6.3[9], while the online system was developed using the tool Eclipse PHP version 3.7.0[10]. BMA is hosted in a Linux Server debian distribution version 8.4, which includes Apache, MySql, and phpMyAdmin for the right operation of BMA.

All software, frameworks, and libraries used in the design and development of BMA have a GNU General Public License (GNU GPL)[11], which implies that BMA was completely created using free software.

We used the nucleotide sequence of genes of different microorganisms extracted from GenBank[12] as a reference sequences. A compilation of resistance mutations described in the literature for genes of different microorganisms, were used for computing the number and type of amino acid variants at the corresponding positions associated with resistance of drugs.

3. RESULTS

The BMA's core is the analysis algorithm that is able to evaluate multiple sequences of different patients, where each patient can include multiple sequences. In addition, the algorithm can analyze desired positions that the analyst can define. The execution of the algorithm is just one part of the complete analysis process. The analysis process includes the following steps:

1. The analyst accesses BMA via the web site and selects the microorganims from the "Mutation Analysis" menu. BMA presents the list of microorganims available, which includes the name, description, and reference sequence (by clicking on the corresponding icon). Figure 1 presents the list of available microorganims.

[1] https://secure.php.net/
[2] http://www.apache.org/
[3] http://getbootstrap.com/
[4] https://jquery.com/
[5] https://github.com/rebuy-de/ezpdf
[6] https://d3js.org/
[7] https://www.mysql.com/
[8] https://www.phpmyadmin.net/
[9] https://www.mysql.com/products/workbench/
[10] https://eclipse.org/pdt/
[11] http://www.gnu.org/licenses/licenses.en.html
[12] www.ncbi.nlm.nih.gov/genbank/

2. The analyst can use the search icon placed in each gene of microorganism to proceed to the following step, which corresponds to the selection of the positions to be analyzed.
3. After selecting the positions to perform the analysis, the analyst is asked to provide the patient sequences as plain text files. BMA can automatically read and analyze multiple data files sequentially. BMA can recognize plain text files, but they have to follow a specific format (see Figure 2). Files must include the symbol '>' and the sequence name in the first line of the file. The sequence data starts on the second line. Nucleotide data must be written in one line. The sequence must include the symbols: A, C, G, T. Sequences can also include the symbol '-' for specifying missing data. In sequences, blank spaces, tabs, break lines and other symbols are not accepted.
4. Once patient files are selected, the analysis algorithm is executed. The algorithm presents the results in three different ways:
    a. Online textual visualization of necessary nucleotide changes that produce an amino acid change, which generates resistance (Figure 3).
    b. An automatically generated report, which is sent to the analyst's e-mail address. This report contains a summary of the calculated mutations for each sequence and the full detailed report of the executed analysis.
    c. A "force-directed" graph that identifies mutations of each patient sequence through node grouping, which corresponds to each analyzed sequence (Figure 4).

The analysis algorithm is based on multiple interactions. It collects all patients' plain text files and iterates in order to analyze all of them independently. For each plain text file, the algorithm collects all sequences. Later on, for each sequence, the algorithm performs a new iteration using the selected positions. Then, for each position, it compares the nucleotide and amino acid of the iterated patient sequence with the reference sequence in the iterated position. At this stage, the information about changes is collected with the corresponding patient, sequence, and position. By finishing the execution of the algorithm, BMA uses the collected results to provide the three aforementioned visualizations.

BMA offers the advantage of sequences analysis obtained by conventional direct sequencing. Furthermore, it accepts relatively long sequences that facilitate the characterization of the resistance mutations. In addition, the analysis for each sequence is done in less than one second.

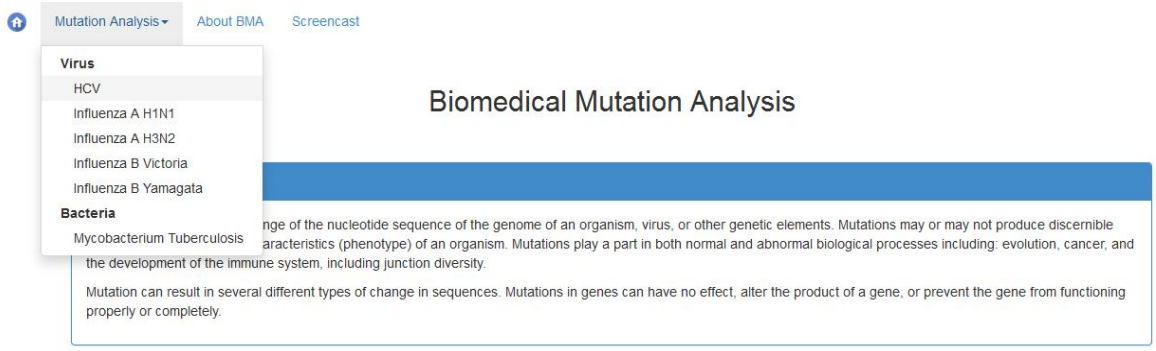

Figure 1. List of available microorganisms.

```
>Patient1
GCGCCTATCACGGCCTACGCCCAACAGACGCGGGGCCTACTTGGCTGCATCATCACCAGCCTCACAGGTCGGGACAAGAACCAGGTCGAGGGAGAGGTTCAAGTGGT
CTCCACTGCAACACAATCTTTCCTGGCGACCTGTGTCAACGGCGTGTGTTGGACTGTTTTCCACGGCGCCGGCTCTAAGACCCTGGCCGGCTCAAAAGGCCCAATCA
CTCAAATGTACACCAATGTAGATCAAGACCTCGTCGGTTGGCAGGCCCCTCCAGGGGCGCGTTCTTTGACACCATGCACCTGTGGTAGCTCAGACCTTTACTTGGTC
ACGAGGCATGCTGATGTCATCCCGGTACGCCGGCGAGGCGACAGCAGGGGGAGCCTGCTCTCCCCCAGGCCTGTCTCCTACTTGAAGGGCTCTTCGGGCGGTCCACT
GCTCTGCCCTTCGGGGCACGCCGTAGGCATCTTCCGAGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATACCCGTTGAGTCTATGGAAACTACCA
TGCGGTCC
>sequence1
GCGCCTATCACGGCCTACGCCCAACAGACGCGGGGCCTACTTGGCTGCATCATCACCAGCCTCACAGGTCGGGACAAGAACCAGGTCGAGGGAGAGGTTCAAGTGGT
CTCCACTGCAACACAATCTTTCCTGGCGACCTGTGTCAACGGCGTGTGTTGGACTGTTTTCCACGGCGCCGGCTCTAAGACCCTGGCCGGCCCAAAAGGCCCAATCA
CTCAAATGTACACCAATGTAGATCAAGACCTCGTCGGTTGGCAGGCCCCTCCAGGGGCGCGGTCTTTGACACCATGCACCTGTGGTAGCTCAGACCTTTACTTGGTC
ACGAGGCATGCTGATGTCATCCCGGTACGCCGGCGAGGCGACAGCAGGGGGAGCCTGCTCTCCCCCAGGCCTGTCTCCTACTTGAAGGGCTCTTCGGGCGGTCCACT
GCTCTGCCCATCGGGGCACGCCGTAGGCATCTTCCGAGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATACCCGTTGAGTCTATGGAAACTACCA
TGCGGTCC
>sequence2
GCGCCTATCACGGCCTACGCCCAACAGACGCGGGGCCTATTTGGCTGCATCATCACCAGCCTCACAGGTCGGGACAAGAACCAGGTCGAGGGAGAGGTTCAAGTGGT
CTCCACTGCAACACAATCTTTCCTGGCGACCTGTGTCAACGGCGTGTGTTGGACTGTTTTCCACGGCGCCGGCTCTAAGACCCTGGCCGGCTCAAAAGGCCCAATCA
CTCAAATGTACACCAACGTAGATCAAGACCTCGTCGGTTGGCAGGCCCCTCCAGGGGCGCGTTCTTTGACACCATGCACCTGTGGTAGCTCAGACCTTTACTTGGTC
ACGAGGCATGCTGATGTCATCCCGGTACGCCGGCGAGGCGACAGCAGGGGGAGCCTGCTCTCCCCCAGGCCTGTCTCCTACTTGAAGGGCTCTTCGGGCGGTCCACT
GCTCTGCCCTTCGGGGCACGCCGTAGGCATCTTCCGAGCTGCTGTGTGCACCCGGGGGGTTGCGAAGGCGGTGGACTTCATACCCGTTGAGTCTATGGAAACTACCA
TGCGGTCC
```

Figure 2. Sequence file format.

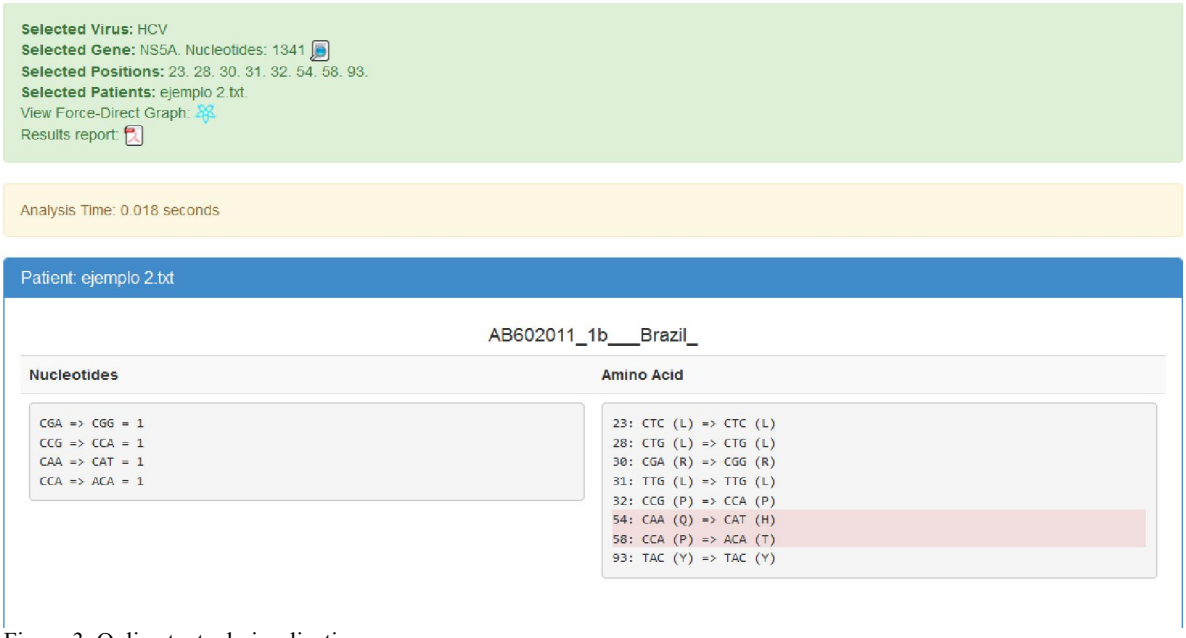

Figure 3. Online textual visualization.

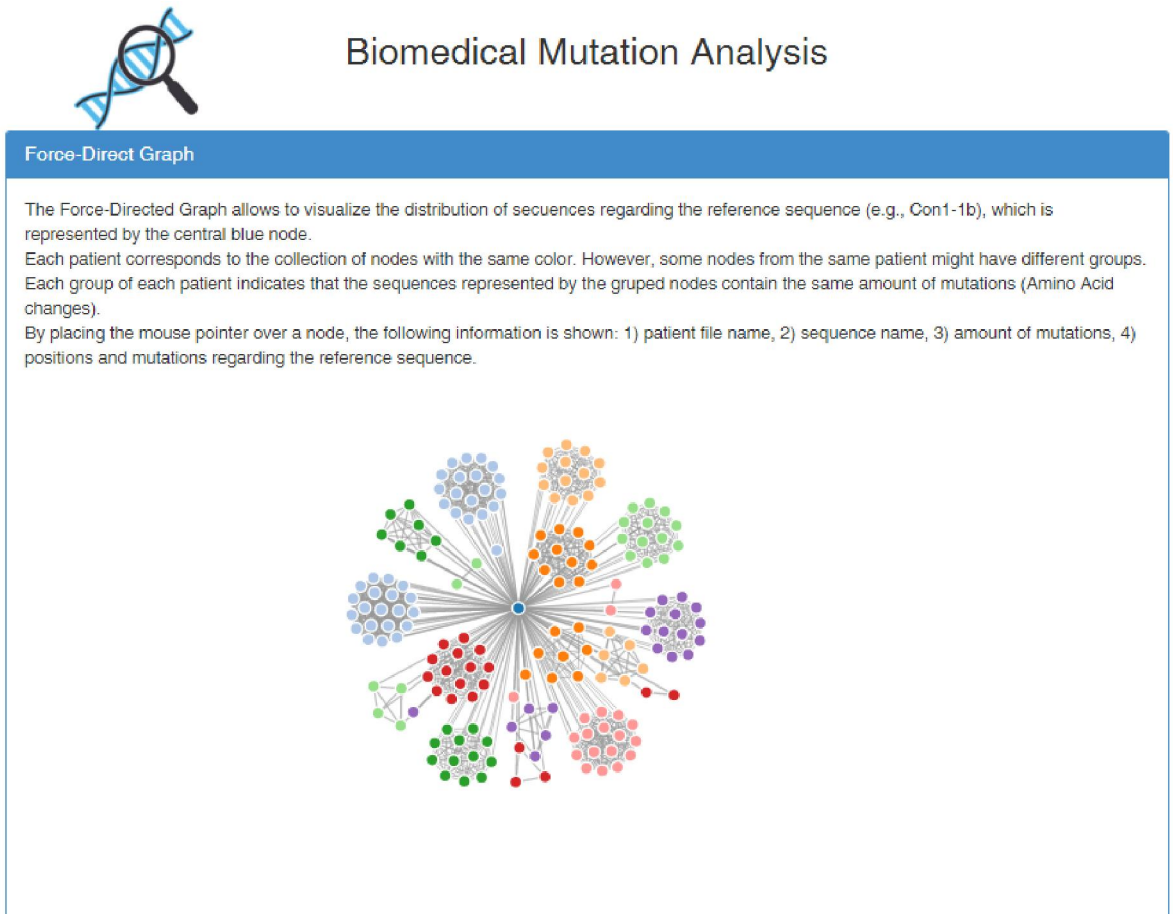

Figure 4. Force Directed Graph visualization.

4. CONCLUSIONS

Software for the detection of mutations associated with resistance to drugs is an important tool, because it guarantees accurate and reliable results. This work provided a software tool called BMA for the detection of drug-resistance mutations in patients' sequences with viral or bacterial infection. BMA is freely available, because it not only allows researchers to perform analysis for the identification of mutations, but also provides detailed information of mutations positions, amino acid changes as well as drugs information and related literature of resistance mutations to the drugs. BMA it provides details of the nucleotides changes that produce an amino acid change.

We obtained an online information system "BMA" that was designed and developed, for performing mutation analysis. BMA provides a suitable analysis facilitating all data management. The results can be visualized in a text report as well as graphically.

BMA provides a quick, easy, and effective computer-based analysis of mutations, including complete documentation and examples. Furthermore, the development of different visualization techniques allows for proper interpretation and understanding of the results. BMA offers a variety of options for reporting sequencing analysis. The results have been generated reporting all specific types of detected variations and providing detailed information about each mutation.

The data obtained by BMA will be useful for the assessment and surveillance of resistance to drugs, avoiding unnecessary treatment failures. BMA may be used by researchers and may help physicians in developing personalized treatment schedules.

For the future work, we also see a potential in adding others microorganisms. We are committed to continuously updating BMA, when novel drugs or resistance patterns are available.

BIBLIOGRAPHY

[1] Miller BF, Furano AV Chen J, "Repair of naturally occurring mismatches can induce mutations in flanking DNA," *Elife*, vol. 3, 2014.

[2] McVey M Rodgers K, "Error-Prone Repair of DNA Double-Strand Breaks," *J. Cell. Physiol*, vol.

1, no. 321, pp. 15-24, 2016.

[3] John S Bertram, "The molecular biology of cancer," *Molecular Aspects of Medicine*, vol. 21, no. 6, pp. 167-223, 2000.

[4] Vincent Burrus and Matthew K Waldor, "Shaping bacterial genomes with integrative and conjugative elements," *Research in Microbiology*, vol. 5, no. 155, pp. 376-386, 2004.

[5] Gelbart WM, Miller JH, et al Griffiths AJF, *Modern Genetic Analysis*.

[6] Bhutta Z, Duse A, et al Laxminarayan R, "The International Bank for Reconstruction and Development / The World Bank," in *Disease Control Priorities in Developing Countries*., 2006.

[7] Chou S Strasfeld L, "Antiviral Drug Resistance: Mechanisms and Clinical Implications," *Infectious disease clinics of North America*, vol. 24, no. 2, pp. 413-437, 2010.

[8] O' Driscoll A, Sleator RD Daugelaite J, "An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics," *ISRN Biomathematics*, vol. 1, no. 14, 2013.

[9] Cho IH, Jeong BC, Lee SH Lee C-R, "Strategies to Minimize Antibiotic Resistance," *International Journal of Environmental Research and Public Health*, vol. 10, no. 9, 2013.

[10] J Rumbaugh, I Jacobson, and G Booch, *The Unified Modeling Language Reference Manual.*: Pearson Higher Education, 2004.