

Store Data from Experiments with Microorganisms Used in Food Industry

Lena Kostadinova-Georgieva^a, Atanaska Bosakova-Ardenska^{*a}, Iliyan Dobrev^b

Received 08th July 2013, Accepted 16th July 2013

Abstract: The aim of this paper is to present results from collaboration of computer engineers and experimenters in microbiology working with molecular-genetic methods. The experimenters in microbiological laboratory at the University of Food Technologies use ARDRA (Amplified Ribosomal DNA Restriction Analysis) analyses and DNA sequencing processed with BLAST (Basic Local Alignment Search Tool) algorithm to identify some microorganisms. Their results have been accumulated in designed database. This will improve the effectiveness and productivity of the molecular-genetic analyses in department of microbiology.

Keywords: DNA analysis, DNA sequence, ARDRA, Database.

1. Introduction

Microbiological analysis of food can be conducted through different methods and the most advanced ones are the molecular-genetic methods. One of them is sequencing a DNA region or gene of a given microorganism, which is subsequently compared to the DNA of reference microbial culture [1]. The sequence is a series of the letters A, C, G and T, which represents the four nucleotide bases in the DNA chain – adenine, cytosine, guanine and thymine. The letters are usually presented next to each other without intervals, e.g. the sequence AAAGTCTGAC. Each series of nucleotides larger than or equal to 4 can be called a sequence. It can be forward or reverse according to its origin form the DNA molecule. The results from the analysis of a sequence are stored in several files of different type. Usually the file, containing the sequence itself is a text file and his name contains information whether the sequence is forward or reverses (R or 27F).

Another commonly applied method for molecular-genetic analysis is ARDRA (Amplified Ribosomal DNA Restriction Analysis). It employs enzyme multiplication of the gene coding the 16S region of the ribosomal DNA and its enzyme restriction. As a result a strain specific ARDRA profile for each microorganism is derived. It includes information about the number of nucleotide bases, contained in the obtained through restriction DNA sequences.

2. Microbiological Analysis

2.1. DNA sequencing and BLAST

BLAST (Basic Local Alignment Search Tool) [2, 3] is an algorithm for finding regions of similarity between biological sequences.

^a University of Food Technologies, department of Computer Systems and Technologies, Plovdiv, Bulgaria

^b University of Food Technologies, department of Microbiology, Plovdiv, Bulgaria

* Corresponding Author: Email: abosakova@yahoo.com;
Phone.: +35932603860

Using BLAST in GenBank [3] the explorers receives four files. For example if we have a sequence with 1102 bases then we receive one text file with bases, one pdf file with signals of different bases in DNA sequence and two files which can be used for visualization of same information in web system. On figure 1 is shown a part of pdf file of one experiment. Using BLAST shows that sequence is recognized as part of genome of microorganism *Lactobacillus casei*.



Figure 2. A transmission line of a power system network.

2.2. ARDRA analysis

The ARDRA analysis is very perspective method for identifying microorganisms. Many researchers works in this field [4, 5, 6]. But now it is not known database which allows thru the internet use ARDRA results to helps some experiments. Both analysis – DNA sequencing processed with BLAST and ARDRA is very useful like express methods for identification of microorganisms in food industry and other fields of human activities. Figure 2 shows results from ARDRA analysis [4]. For example *Lactobacillus*

helveticus (number 16) has a profile with three parts with sizes 325, 194 and 119 bases and Lactobacillus plantarum (number 17) has a profile with two parts sized 300 and 259 bases.

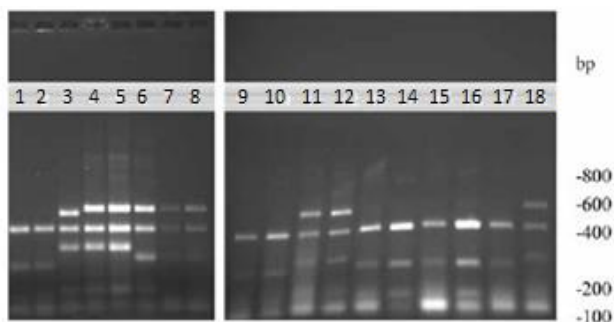


Figure 2. ARDRA profile of Lactobacillus strains obtained by digestion of amplified 16srDNA with HaeIII

3. Aim

The accumulated results from sequence analyses in the department of Microbiology at the University of Food Technologies were obtained through the application of the algorithm for sequence comparison BLAST. The aim of the designed database is to facilitate the organization and interpretation of this data, as well as the results from ARDRA analyses. This will improve the effectiveness and productivity of the molecular-genetic analyses for microbiological applications.

4. Design of Database

4.1. Design of ER model

When designing a database so-called ER model (Entity Relationships Model) is used. After analyzing the work in the laboratory the following objects have been identified:

- 1) Experiment is carried out by a particular experimenter at a specified date and is characterized by a short description and name (label);
- 2) ARDRA analysis: analysis that can be performed as part of the work on an experiment and give information about profile obtained through the use of a specific restriction enzyme;
- 3) Operator: experimenter who has a name and occupies a specific position;

- 4) Sequence (DNA sequence): naming sequence of nucleotide bases (part of the genome of the organism) received on a certain date and stored in a file;
- 5) Recognition: probability of identifying the received sequence as part of genome of a known microorganism.

Figure 3 shows the designed database ER model.

Objects shown in Figure 3 are related as follows:

- The experiment is conducted by an operator and may include results of ARDRA analysis and/or DNA sequences of organisms studied;
- The operator can perform experiments and occupies a specific position;
- ARDRA analysis is part of a concrete experiment, and the result of using specific restriction enzymes;
- The sequence was obtained in a concrete experiment, a type of segment of DNA, which is extracted and can be recognized as part of the genome of some organisms;
- Recognition is performed for a particular sequence by using an algorithm for detection in an existing genetic base and to show what is the probability the test sequence to be part of the genome of a known microorganism.

4.2. Design of Relational model

ER model is transformed into a relational model [7], consisting of a set of tables (relations) and relations between them. Figure 4 shows the structure of the relational database model. As DBMS (Database management system) is used MS Access 2007. The database is normalized to the first and second normal form [7]. The third normal form is not applied. In table Sekvencii we have two columns for which the values are connected. These columns are: Sequence and NumBase. The length of string in column Sequence is equal to number in column NumBase. But this is necessary for controlling correct input of sequences. We can check whether length of string in Sequence is equal to number in NumBase. If this is not true then we will search correct information in source files.

5. User Interface

Along with designing the database a graphical user interface (GUI) with tools provided by MS Access 2007 is also designed. Figure 5 shows the list of the developed forms.

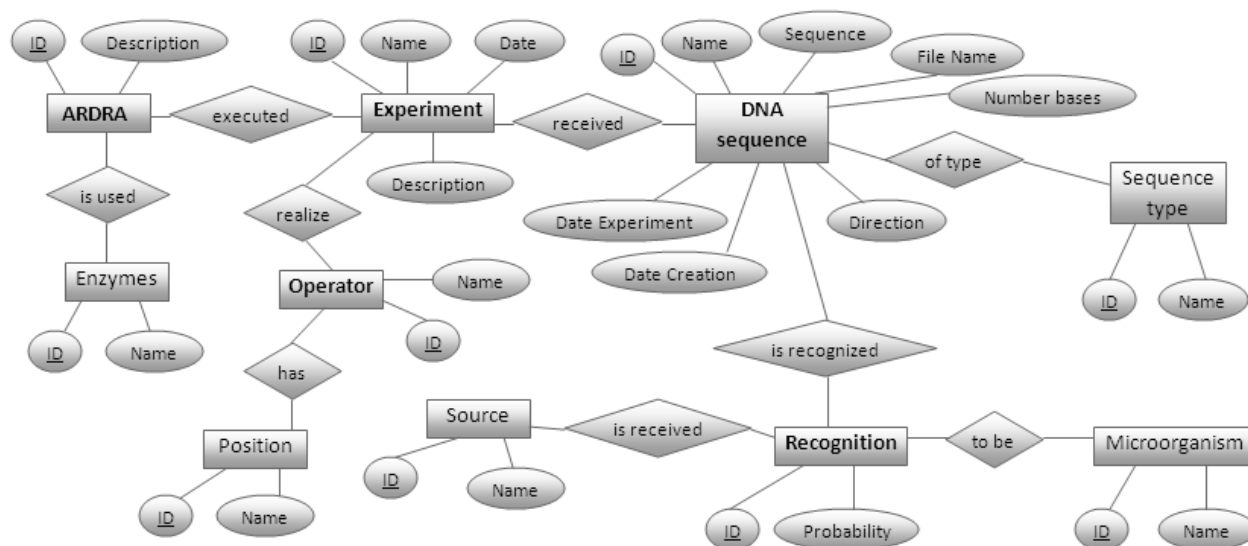


Figure 3. Entity Relationships model

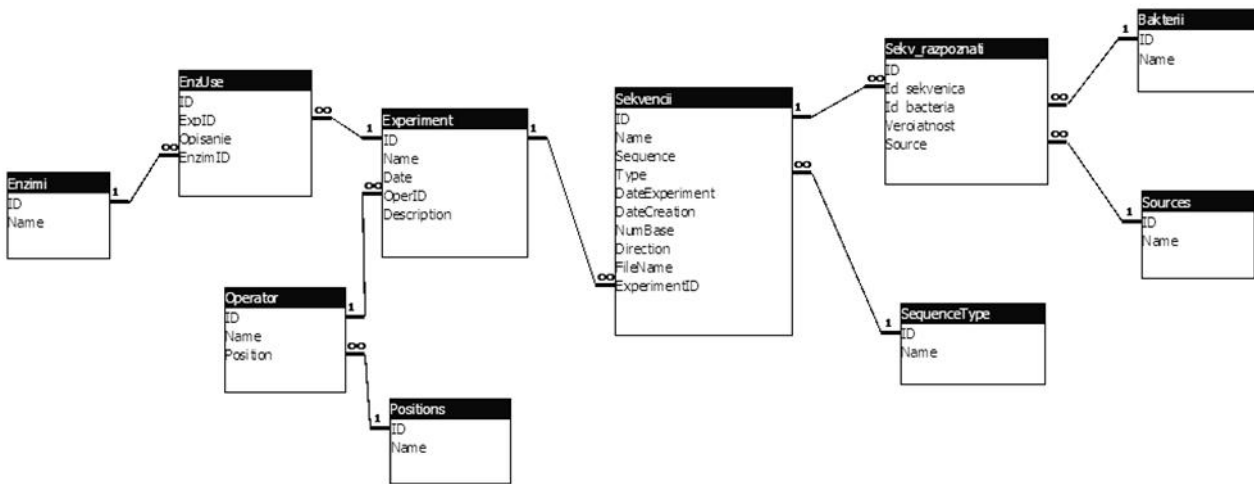


Figure 4. Relational model

When one opens the ACCDE a database file is loaded in a form Home. It contains a list of operators and after being selected an operator may consider information about conducted experiments (Fig. 6). Each experiment has a name, description and start date.

- Information about selected by user sequence;
- ARDRA analysis for a selected experiment;
- Recognitions that are made for a specific sequence.



Figure 5. GUI - List of forms

Figure 7. Form "Add DNA Sequence"

Name	Date	Description
Experiment Rossi	20.2.2013 r.	--
Experiment Alexander	21.2.2013 r.	--
Second experiment	21.2.2013 r.	This is the next step in our research
Third exp	28.3.2013 r.	

Figure 6. Form "Experiments"

For each experiment can be viewed and edited as DNA sequences also ARDRA analyzes. Figure 7 shows a form for adding sequences.

The sequence can be recognized as part of the genome of a known microorganism and this information is being edited through the forms shown in Figure 8.

Reports have been created to be used for printing the following views:

- Experiments made by the selected operator;
- Sequences associated with specific experiment;

Microorganisms name	Probability	Source
Lactobacillus hayakitensis 16S ribosomal RNA, partial sequence	99	http://blast.ncbi.nlm.nih.gov/
Lactobacillus delbrueckii subsp. lactis DSM 20072 strain ATCC 12315	99	http://blast.ncbi.nlm.nih.gov/
Lactobacillus delbrueckii subsp. delbrueckii strain BCR12195 16S ribosomal	99	http://blast.ncbi.nlm.nih.gov/
Lactobacillus acidophilus strain BCR10695 16S ribosomal RNA, complete se	95	http://blast.ncbi.nlm.nih.gov/
Lactobacillus formicilis strain TV1018 16S ribosomal RNA, partial sequence	95	http://blast.ncbi.nlm.nih.gov/
Lactobacillus gallinarum strain ATCC 33199 16S ribosomal RNA, partial sequ	95	http://blast.ncbi.nlm.nih.gov/
Lactobacillus kalixensis strain Kx12742 16S ribosomal RNA, complete sequ	95	http://blast.ncbi.nlm.nih.gov/
Lactobacillus helveticus DSM 20075 16S ribosomal RNA, complete sequence	95	http://blast.ncbi.nlm.nih.gov/
Lactobacillus litospatensis strain JCM 1039 16S ribosomal RNA, partial sequen	95	http://blast.ncbi.nlm.nih.gov/
Lactobacillus crispatus strain ATCC33820 16S ribosomal RNA, partial sequen	95	http://blast.ncbi.nlm.nih.gov/

Figure 8. Forms "Identified microorganisms" and "Recognize sequence"

The advantage of the created database is that it allows aggregation of results of analysis of sequences obtained using different algorithms [8].

In addition, for each experiment can be performed and ARDRA analysis that complement the information obtained by using algorithms for comparing in the existing genetic databases. Figure 9 shows report for ARDRA results of one experiment.

ARDRA

Experiment: Experiment Rosi

Description	Enzymes
using ARDRA - 400/500 kB	enzim
400/300 kB	enzim
using ARDRA analysis receive: 300/700 kB	enzim 2
345/120 s korekcii 20/35	enzim 3
345/250 s korekcii 20/35	enzim 3

Figure 9. Report of ARDRA results

6. Conclusion And Future Work

A database for the microbiological laboratory has been designed and implemented. The structure of the database enables to store information for identification of microorganisms obtained from various sources, and also the results from ARDRA analysis. This will significantly increase the efficiency and productivity of work with molecular genetic techniques used in microbiological practice. The work of the team continues in the following areas:

- improving the user interface of the database;
- accumulation of experimental data;

- development of an effective algorithm for detection of DNA sequences;
- publishing database in Internet.

References

- [1] I. Murgov, Z. Denkova, Genetics and selection of microorganisms, UFT Academic Publishing House, Plovdiv, 2007 (in Bulgarian)
- [2] <http://en.wikipedia.org/wiki/BLAST>
- [3] <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- [4] T. Stefanova, Urshev Z., Dimitrov Z., Fatchikova N., Minkova S. Species affiliation of dairy lactobacilli with angiotensin converting enzyme inhibitory activity. *Biotechnol. & Biotechnol. Eq.* 23, pp. 1250-1254, 2009.
- [5] I. Dobrev, Application of lactobacilli and bifidobacteria in the production of healthy foods from goat milk, PhD thesis, University of Food Technologies, Plovdiv, 2011
- [6] Petrov K., Urshev Z., Petrova P. L(+)-Lactic acid production from starch by a novel amylolytic *Lactococcus lactis* subsp. *lactis* B84. *Food Microbiology*, vol. 25, pp. 550-557, 2008.
- [7] Michael J. Hernandez, Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design, Addison-Wesley Professional, 2003
- [8] Li L, Khuri S. A comparison of DNA fragment assembly algorithms. In: International conference on mathematics and engineering techniques in medicine and biological sciences; pp. 329-35, 2004.