

A review

An Introduction to Bioinformatics

VLAIC Bogdan*

Faculty of Animal Husbandry and Biotechnology, University of Agricultural Sciences and Veterinary Medicine Cluj - Napoca, Mănăştur St., No. 3 – 5, 400327 Cluj-Napoca, Romania

Received 2 February 2015; received and revised form 25 February 2015; accepted 27 February 2015
Available online 1 September 2015

Abstract

The paper aims to emphasize general aspects concerning bioinformatics object of activity, focusing on definitions, and main areas where this complex issue has uses. Being an interdisciplinary field of science, it involves biology, molecular biology knowledge, as well as computational skills. Mastering techniques specific for bioinformatics, allows researcher to perform large scale comprehensive analysis on a given subject. Bioinformatics has an innovative contribution to identification of deepest insights of the living cells, in the meantime with generating the whole picture of the interesting cell, and for this reason it has to develop in the same rhythm with science fields it is involved. In this respect, the paper also presents the main challenges that bioinformatics must face in the future.

Keywords: protein, DNA, RNA, database, sequences.

1. Background

The simultaneous development of genetics as biological science and computer science led, by the end of '80s to a common point between these science branches. In this way, resulted the interdisciplinary scientific tool, known ourdays as bioinformatics [6].

Literature cites a series of definitions of above mentioned research field, and among these we mention "a union of biology and informatics" [5], or, according to English Oxford Dictionary, a more detailed definition focuses bioinformatics as application of computer and statistics sciences in biology and/or biological related fields, as a management information system, in order to get full understanding of information associated to biological field conceptualized in terms of molecules [9].

Bioinformatics as independent scientific tool, is a term similar to computational biology, even their sense is not the same [4]. Bioinformatics denominate a series of activities of database type, which aim to obtain and maintain data sets in a safe state for indefinite time periods, while computational biology denominate the use of specific algorithms in order to perform punctual biological analyses [3].

Both these scientific branches located at the interface between biology and informatics/computer science, are based on a specific infrastructure that includes the entire management system of all kind of specific information, including analysis tools [6].

2. Area of interest

Taking into consideration the above mentioned definitions and describing of what bioinformatics concept expresses [2, 5], it is obvious that it is, above all, a complex interdisciplinary issue, which involves several separate scientific domains (Fig. 1).

* Corresponding author.
Tel: +40-264-596384
Fax: +40-264-593792
e-mail: bogdan.vlaic@yahoo.com

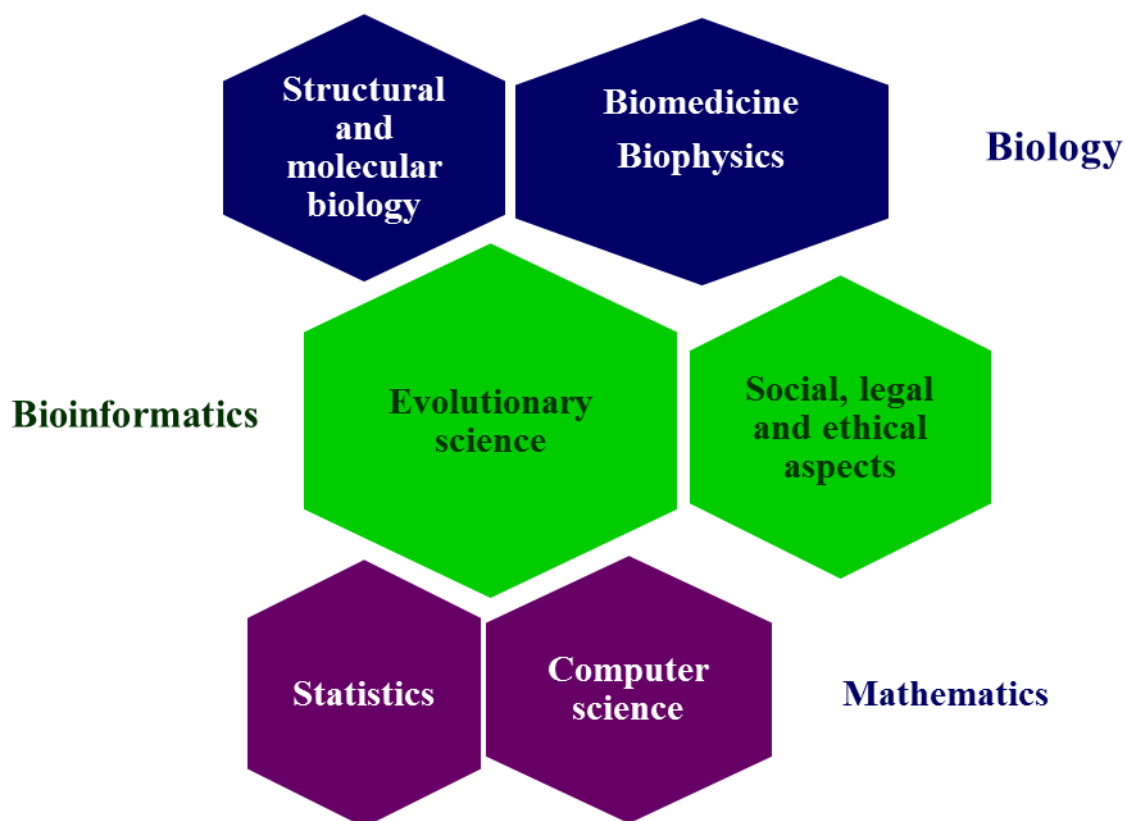


Figure 1. The interdisciplinary fields of bioinformatics

Bioinformatics, as scientific interdisciplinary tool, acts on three plans (Fig. 2): organization of different biological data, development of both resources and instruments in

order to facilitate data analysis, and, also using available instrumental portfolio for analysing data and as supplementary task, their appropriate interpretation [2].

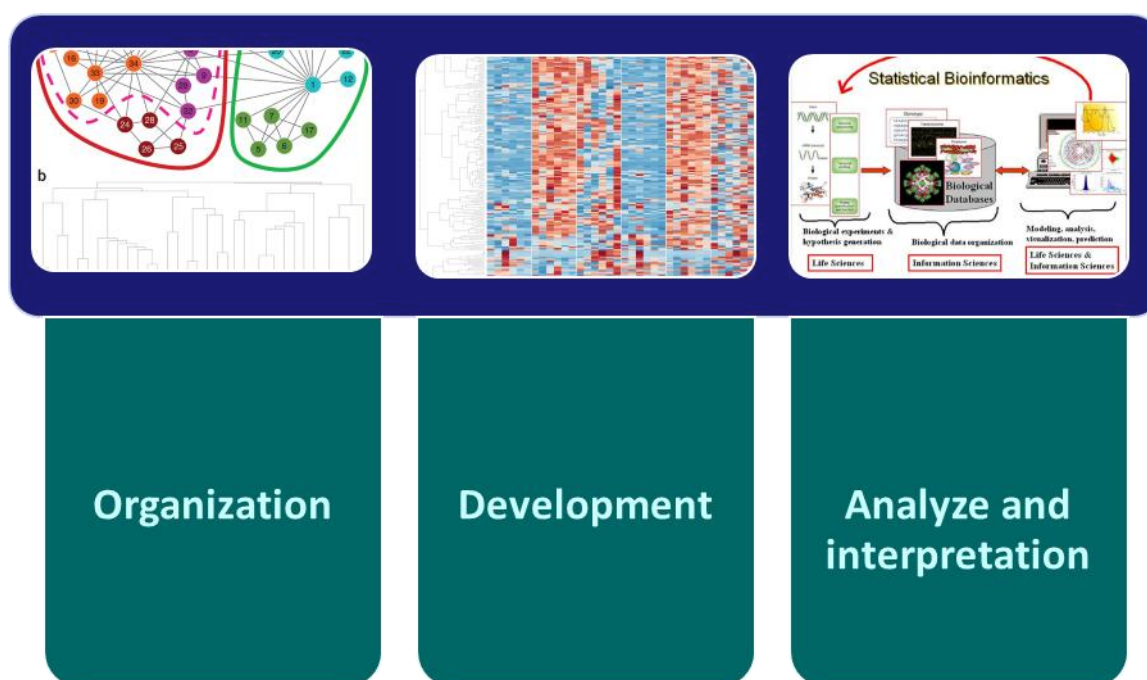


Figure 2. The bioinformatics applications [11, 12, 13]

The organization of data is both meant to facilitate researchers to access available information, and also supplying possibility for researchers to perform new entries, according to

their research results. As examples, we mention protein data banks for 3D macromolecular structures, as RCSB PDB (Fig. 3a) [14], or Protein Data Bank Archive (Fig 3.b) [1, 15].

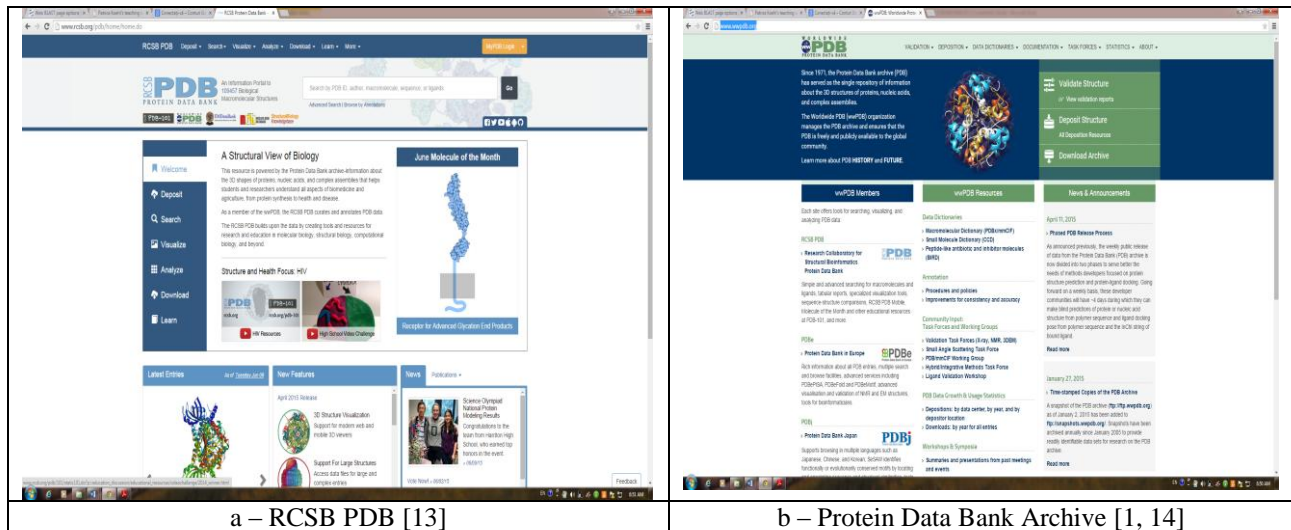


Figure 3. Examples of databases

The development of specific tools within bioinformatics area concerns mainly sequencing area. This operation is made in order to perform comparisons between different proteic or DNA fragments with other previously available, aiming to

find possible matches with significance from biological point of view.

In this respect, a series of specific databases are used – e.g. PSI-BLAST (Fig. 4) [16],

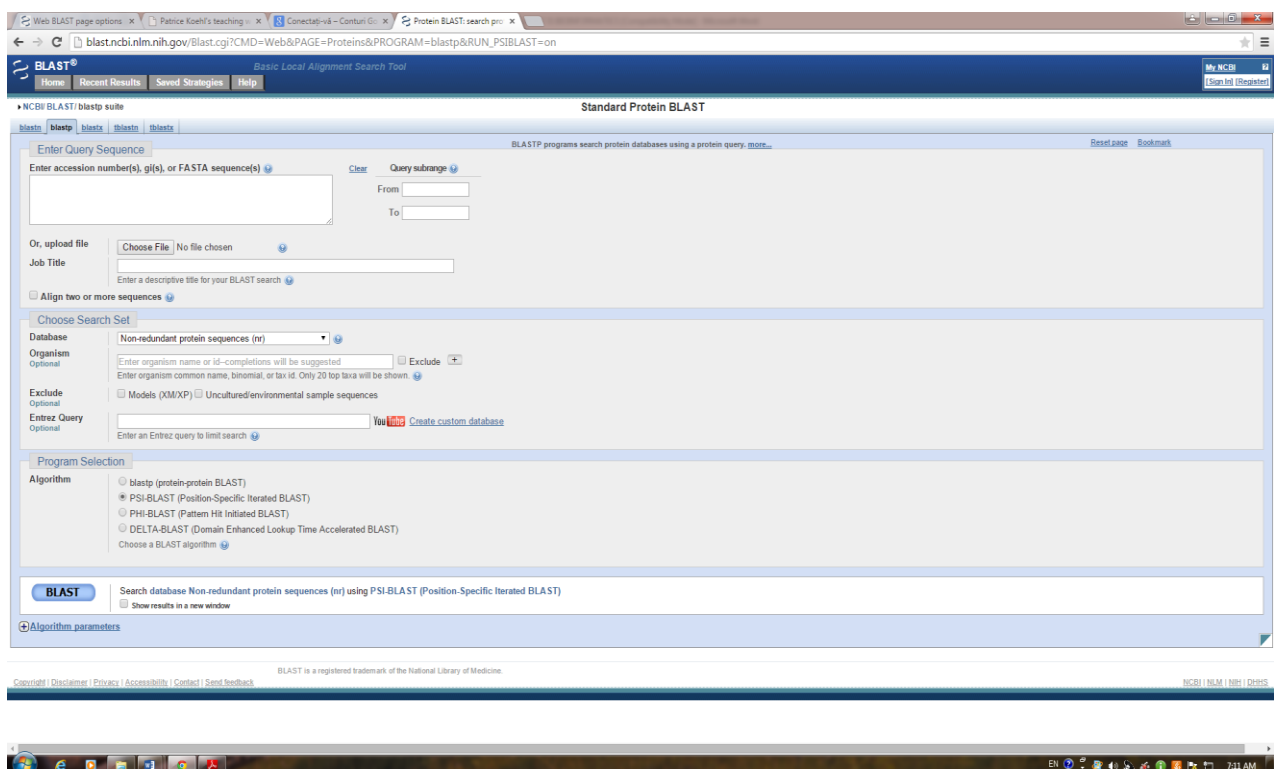


Figure 5. Example of sequence in PSI-BLAST databases [16]

or FASTA [9] (Fig. 5),

```
>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDLTLVLVNAIYFKGMWKTAFAEDTREMPFHVTKQESKPVQ
MMCMNNSFNVAITLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKL
TEWTNPNTMEKRRVKVYLPQMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSA
ESLKISQAVHGAFMELEDGIEMAGSTGVIEDIKHSPSESEQFRADHPFLFIKHNP
TNTIVYFGRYWSP
```

Figure 5. Example of sequence in FASTA format [10]

If great importance in this context, is the fact that the used must have important expertise in both molecular biology and computational techniques.

Mastering techniques specific for bioinformatics, allows researcher to perform large scale comprehensive analysis on a given subject.

Thus, bioinformatics supplies important tools in order to acquire better knowledge of living organisms in terms of intimate molecular structures, and their functionality. It also has an innovative contribution to identification of deepest insights of the living cells, in the meantime with generating the whole picture of the interesting cell [7].

3. Concluding remarks

Our days, when scientists all over the world, from all research areas, but especially those involved in life sciences domains confronts with an increasing amount of data of all kinds, computational methods are of most great importance. Thus, the use of bioinformatics must align to rapid developments in the field of biology.

According to Birney, Burge and Fickett, cited by Koehel, in near future, the use of bioinformatics as tool in development of life sciences is challenged by a series of outputs that impose special performances and accomplishing several scientific needs [8]. Thus, bioinformatics must face the following issues, which involve the necessity of deliver:

- more accurate models in order to predict where and when transcription will occur in a given genome (initiation and termination of the process), in the meantime with improvement of capacity of prediction concerning occurrence of a precise transcription within a genome.
- tools for accurate predictive models of alternative RNA splicing. In this way, the researcher will gain improved capacity to

- predict the splicing pattern of any primary transcript in any tissue.
- accurate models of signal transduction pathways. With these tools, the researcher will have methodological portfolio to predict cellular responses to external stimuli
- accurate tools in order to allow determination of recognition codes for protein : DNA, protein : RNA, protein : protein.
- accurate tools for accurate ab-initio protein structure prediction.
- accurate tools for rational design of small molecule inhibitors of proteins.
- accurate tools for understanding of protein evolution, at intimate level (mechanistic), meaning understanding exactly how new protein functions evolve.
- accurate tools for understanding of speciation (mechanistic), meaning ability to deliver molecular details of how speciation occurs.
- accurate tools for development of effective gene ontologies. In this way, there will be possible to identify systematic ways to describe gene and protein function.
- opportunities to be included in educational programmes, by development of bioinformatics curricula.

References

- [1] Bernstein F.C., T.F. Koetzle, G.J. Williams, E.F. Meyer, M.D. Jr., Brice, J.R. Rodgers, 1977, The Protein Data Bank. A computer-based archival file for macromolecular structures, Eur J Biochem, 80 (2), 319-24.
- [2] Gerstein M., 2000, Integrative database analysis in structural genomics. Nature Struct. Biol., 7, 960-963.
- [3] Hofestädt R., T. Lengauer, M. Löffler, D. Schomburg (Eds.), 1997, Bioinformatics, German Conference on Bioinformatics, Leipzig, Germany, Springer, ISBN 3-540-63370-7

- [4] Lim H.A., and C.R. Cantor Eds., 1995, Bioinformatics and Genome Research, World Scientific Pub. Co., New Jersey.
- [5] Luscombe N.M., S.E. Austin, H.M. Berman, J.M. Thornton, 2000, An overview of the structures of protein-DNA complexes, *Genome Biology*; 1(1), 1-37.
- [6] Venkatesh T.V., B. Bowen, and H.A. Lim, 1999, Bioinformatics, pharma and farmers, *Trends in Biotechnology*, 17 (3), 85-88.
- [7] Wilson CA, J. Kreychman, M. Gerstein, 2000, Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. *J. Mol Biol*, 297 (1), 233-49.
- [8] ***, Koehl P., Computational Structural Bioinformatics, http://www.dtrends.com/Bioinformatics/koehl_UCDavies.html
- [9] <http://www.ncbi.nlm.nih.gov/BLAST/blastcgihelp.shtml>
- [10] [https://www.google.ro/search?q=development +of+computational&biw=1091&bih=918&source](https://www.google.ro/search?q=development+of+computational&biw=1091&bih=918&source)
- [11] [https://www.google.ro/search?q=development +of+computational+tools&biw=1091&bih](https://www.google.ro/search?q=development+of+computational+tools&biw=1091&bih)
- [12] [https://www.google.ro/search?q=organization+of +biological+data](https://www.google.ro/search?q=organization+of+biological+data)
- [13] www.rcsb.org
- [14] <http://www ww pdb.org/>
- [15] ***, English Oxford Dictionary, <http://www.thefreedictionary.com/Oxford+English+Dictionary>
- [16] http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE=Proteins&PROGRAM=blastp&RUN_PSIBLAST=on