TEHNIKA I INFORMATIKA U OBRAZOVANJU
4. Internacionalna Konferencija, Tehnički fakultet Čačak, 1–3. jun 2012.
TECHNICS AND INFORMATICS IN EDUCATION
4th International Conference, Technical Faculty Čačak, 1–3rd June 2012.

UDK: 371.3::004 Stručni rad

TECHNICALLY SUPPORTED BIOINFORMATICAL EDUCATION

Done Stojanov[1], *Todor Cekerovski*[2], *Gabriela Suteva*[3]

Summary: *Bioinformatics is a new scientific discipline, connecting together technology and biology. Large amounts of data coming from organism's micro-level can't be processed without computer's support. Nowadays many bioinformatics tools are internet-available. Using these resources for practical instruction in bioinformatics helps students easily understand difficult theoretical concepts, making them being able to access genetic database and to process genetic material using genetic processing software. Bioinformatics curriculum and software tools, being used for practical curriculum improvement at the Faculty of Computer Science are topics of discussion in this paper.*

Keywords: *Bioinformatics, technical tools, curriculum.*

TEHNIČKI PODRŽANO BIOINFORMATIČKO OBRAZOVANJE

Rezime: *Bioinformatika je nova naučna disciplina, povezujući zajedno tehnologije i biologije. Ogromne kolicine podataka, koji dolaze iz organizma na mikro-nivou ne može biti obrađeni bez podrške računara. Danas mnogo bioinformatičke alatke su internet dostupne. Upotrebom ovih sredstava lako je razumjeti komplikovane teorijse koncepte, omogućavati studenata biti u mogućnosti pristupati genetsku bazu podataka i obrađivati genetski materijal, koristeči softver za genetsku obradu. Bioinformatički nastavni plan i softverskih alata, koji se koriste za poboljšanje praktičnog nastavnog programa na Računarskom Fakultetu su teme diskusije u ovom radu.*

Ključne reči: *Bioinformatika, tehničkih alata, nastavni plan.*

1. INTRODUCTION

Decades ago, biology and computers science seemed to have nothing in common. Abrupt technological changes in the previous century denied this hypothesis. Large amount of data obtained from organism's micro-level undergoes computer processing, producing output results, which can serve as indicators for possible evolutionary relationship between the species. Processing genetic material with thousands, in some cases million compartments,

---

[1] Junior Ass. MSc. Done Stojanov, Faculty of Computer Science, University „Goce Delcev" , Krste Misirkov nn, Stip, E-mail: done.stojanov@ugd.edu.mk

[2] Ass. MSc. Todor Cekerovski, Faculty of Computer Science, University „Goce Delcev" , Krste Misirkov nn, Stip, E-mail: tose.cekerovski@ugd.edu.mk

[3] Gabriela Suteva, Faculty of Computer Science, University „Goce Delcev", Krste Misirkov nn, Stip, E-mail: gabriela.suteva@ugd.edu.mk

is time and effort expensive task without assistance of computer software, which may solve the problem in a couple of seconds. Faced with this requirement, many computer scientists working on problems in bioinformatics have directed their efforts in developing time-efficient algorithms, requiring as less as possible memory.

In this paper bioinformatics curriculum at the University „Goce Delcev" – Stip will be presented, emphasizing the use of freely accessible bioinformatics software and resources, as means for improving instruction's quality.

2. CURRICULUM

Bioinformatics curriculum at the University „Goce Delcev'' is organized into ten sections, covering different topics. Prior advanced algorithms introduction, students have to be familiar with the basic biological concepts. Structural and functional aspects of genetic material: DNA, RNA and proteins, are subjects of instruction in the first two weeks. After understanding basic biological concepts, assuming that as computer science students they have already gained mathematical, logical and computer programming skills, the types of DNA alignment techniques and the DOT-matrix method, as basic alignment methodology, are presented to the students. Local and global pairwise alignment techniques are next elaborated, presenting in details the Smith-Waterman method as prime local pairwise alignment technique and the Needleman-Wunsch algorithm for global pairwise alignment.

Smith-Waterman and Needleman-Wunsch algorithm are not genetic database searching techniques. The problem of finding database samples with similar structure regarding referent sequence is solved with heuristically-based applications, such as: FASTA, BLAST and PSI BLAST, being planned for instruction in the fifth week.

Despite the pairwise alignment techniques, there are also multiple alignment techniques, applicable on more than two genetic sequences. Since more than two sequences are subjects of alignment, these alignments are computationally expensive and therefore they are heuristically-based. Multiple alignment methods and applications are presented to the students in the sixth week.

By the end of the seventh week, students will be familiar with the phylogenetic trees and the distant matrix methods, especially with the UPGMA method.

Proteins processing methods are planned for instruction at the end of the semester. Protein Data Bank, methods for comparing and predicting protein 3D structure, knowing protein's primary structure will be presented to the students. Concrete protein 3D visualizing applications are planned to be presented to the students.

Table 1: Bioinformatics curriculum at the University „Goce Delcev"

Week	Topic
1	molecular biology introduction – part 1
2	molecular biology introduction – part 2
3	types of alignment techniques, DOT-MATRIX method
4	local and global alignment techniques, Smith-Waterman and Needleman-Wunsch algorithm
5	genetic database searching techniques: FASTA, BLAST and PSI BLAST
6	multiple sequence alignment
7	phylogenetic trees, distant matrix methods and UPGMA method
8	PROTEIN DATA BANK, structure of .pdf file, protein structure comparison methods
9	protein 3D structure predicting models
10	protein 3D structure visualization methods

3. INSTRUCTIONAL SUPPORT TOOLS

Software tools being used for practical instruction are given in Table 2. Next to the application name, a short tool's description is also given. Students have found these tools as useful and learning facilitating means, successfully connecting the theory and practice.

Table 2: Bioinformatics tools being used in the practical instruction

Tool	Tool's description
DOTTER	*Graphical dot plot program, representing exact matching regions as continuous diagonals of dots.*
EMBOSS Needle	*Creates an optimal global alignment of two sequences, using the Needleman-Wunsch algorithm.*
EMBOSS Water	*Creates an optimal local alignment of two sequences, using the Smith-Waterman algorithm.*
FASTA SERVICE at: http://www.ebi.ac.uk/Tools/sss/fasta/	*Genetic database searching service for similar sequences, using FASTA methodology*
BLAST SEVICE at: http://blast.ncbi.nlm.nih.gov/Blast.cgi	*Genetic database searching service for similar sequences, using BLAST methodology*
ClustalW2, T-Coffee, Muscle	*EBI online tools for multiple sequence alignment*
TRED	*A tool to visualize and edit phylogenetic tree*
RCSB PDB Protein Comparison Tool	*Proteins structure comparison tool*
MODELLER	*A computer tool for producing homology models of protein's tertiary and quaternary structure*
Jmol	*Open-source Java viewer for protein's chemical structure in 3D space.*

4. CONCLUSION

In this paper bioinformatics curriculum at the University „Goce Delcev" have been presented. In order to understand difficult bioinformatics concepts, the use of tools, implementing different algorithms, is more than necessary. Therefore a list of applications, implementing concrete techniques, has been given. Including these applications in the practical instruction, students have found more than useful.

5. REFERENCES

[1] Notherdame C., Higgins L., Heringa J., O'Sullivan O., Suhre K. Abergel C.: *T coffee: Multipurpose sequence alignment program*, Journal of Cell and Molecular Biology 7(2) & 8(1), 2010., pp. 71-72.

[2] Fisher A., Sali A.: *Modeller: generation and refinement of homology-based protein structure models*, Methods in Enzymology 374, 2003., pp. 461-491.

[3] Sonnhammer E., Durbin R.: *A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis*, Gene 167, 1996., pp. GC1-10.

[4] Higgins D., Sharp P.: *CLUSTAL: a package for performing multiple sequence alignment on a microcomputer*, Gene 73, 1988., pp. 237-244.

[5] http://www.ebi.ac.uk/Tools/psa/

[6] http://blast.ncbi.nlm.nih.gov/Blast.cgi

[7] http://code.google.com/p/tred/
[8] http://www.rcsb.org/pdb/workbench/workbench.do
[9] http://jmol.sourceforge.net/