COURSE UNIT DESCRIPTION - BIOINFORMATICS

Course unit title	Code
BIOINFORMATICS	

Lecturer(s)	Department(s)
Coordinator: Assoc. prof. Visvaldas KAIRYS **Other(s):**	Institute of Biotechnology of Vilnius University, V.A.Graičiūno 8, Vilnius LT-02241

Cycle	Level of the course unit	Type of the course unit
Full-time studies (1st stage)	1 out of 1	Obligatory

Mode of delivery	Period of delivered	Language(s) of instruction
Face-to-face	4th semester, spring	Lithuanian (English)

Prerequisites and corequisities	
Prerequisites: Basic knowledge in Biochemistry, Genetics, Statistics	**Corequisities (if any):** Computer literacy

Number of credits allocated to the course unit	Student's total workload	Contact hours	Self-study and research hours
5	134	64	70

Purpose of the course unit: programme competences to be developed

The course unit aims to develop:

Specific competences:
- skills to collect information about biological molecules and to manipulate that information, using available resources and programs;
- understanding of various bioinformatics algorithms, their applicability, drawbacks and limitations

General competences:
- analytical and critical thinking ;
- ability to communicate in written and verbal forms using correct Lithuanian in professional field;
- ability to communicate in written and verbal forms using correct English, to use information sources in English in professional field;
- skills for self-development, learning skills in order to study general science resources;

Learning outcomes of the course unit	Teaching and learning methods	Assessment methods
Ability to use primary gene, mRNR, protein sequence databases such as NCBI and Uniprot.	Lectures, self-directed learning	Midterm
Understanding Needleman-Wunsch and Smith-Waterman sequence alignment algorithms. Capability to use BLAST and other related sequence alignment programs and understand their output.	Lectures, self-directed learning	Midterm
Capability to use advanced sequence alignment programs such as PSI-BLAST, PHI-BLAST and HMMER, and understand the situations where to apply them.	Lectures, self-directed learning	Midterm
Capability to use multiple sequence alignment programs, e.g. ClustalW, and understand their difference from the alignment of two sequences.	Lectures, self-directed learning	Midterm
Understand different phylogenetic tree types and be able to construct them using available programs.	Lectures, self-directed learning	Final exam
Capability to use protein structure visualization	Lectures, self-directed learning	Final exam

program RasMol, and capability to use command line options for full control.		
Capability to perform advanced searches in the RCSB Protein Data Base. Understanding of the PDB file format and its limitations.	Lectures, self-directed learning	Final exam
Capability to use various protein modelling servers and programs, e.g. secondary structure prediction for the unknown structure, sidechain modelling, loop modelling, etc. Understanding difference between homology modelling, threading and ab initio modelling, learning to use 3D homology modelling servers.	Lectures, self-directed learning	Final exam

Content: breakdown of the topics	Contact hours							Self-study work: time and assignments	
	Lectures	Tutorials	Seminars	Exercises	Laboratory work	Internship/work placement	Contact hours	Self-study hours	Assignments
1. Introduction to bioinformatics. Genes, mutations. GenBank and its contents.	1			2			**3**	**3**	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
2. NCBI Entrez search system. RefSeq annotated sequences. Monogenic disorder database OMIM. Primary and Secondary databases. Uniprot protein database. Genome browsers.	1			4			**5**	**5**	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
3. Homology. Protein motifs, domains. Sequence alignment. Alignment quality measures. Identity matrix. Dot matrix method. Moving window method.	1			3			**4**	**4**	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
4. PAM, BLOSUM scoring matrices, their derivation and application.	1			3			**4**	**5**	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
5. Needleman-Wunsch and Smith-Waterman alignemnt algorithms. Gap penalty.	1			3			**4**	**4**	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
6. BLAST alignment and algorithm. Statistical alignment measures (E-score, Z-score, etc.). BLAST parameters and their application.	1			3			**4**	**5**	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
7. PHI-BLAST. PSI-BLAST. Alignment programs and databases which use Markov Chain models.	1			3			**4**	**4**	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.

#	Topic									
8.	Multiple sequence alignemnt program ClustalW, its algorithm. Other multiple sequence alignment programs.	1			3			4	4	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
9.	Philogenetic trees, their types, distance between sequences types, philogenetic tree construction programs.	1			3			4	5	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
10.	Three-dimensional protein structures, their visualization.	1			4			5	5	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
11.	Basics of protein structure. Ramachandran plots. Secondary, ternary, quaternary structures. PDB, mmCIF formats. Protein data bank. Primary and secondary structure databases.	1			2			3	4	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
12.	Accuracy of structural methods. Errors in structure determination, their detection. Protein three-dimensional alignments.	1			3			4	5	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
13.	Structural folds. SCOP, CATH databases. Automatic detection of secondary structure elements. Domain determination.	1			3			4	4	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
14.	Prediction of protein function. Secondary structure prediction. Special aspects of modeling: solvent accessibility, transmembrane domain, etc.	1			3			4	4	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
15.	Homology modeling, threading, ab initio modeling. Loop modeling, sidechain modeling.	1			4			5	5	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
16.	Homology modeling servers. Three dimensional structure prediction competitions (CASP and others). The overview of the latest CASP competition results.	1			2			3	4	Self-directed learning of the topic-related textbook material, analysis of the topic-related scientific papers.
	Total	**16**			**48**			**64**	**70**	

Assessment strategy	Weight,%	Assessment period	Assessment criteria
Homework assignments	25	After the exercise sessions	Evaluated for accuracy and completeness using 10 pt system. As as additional criterion, every day for turning the due assignment late is penalized by 10%.
Presentation of an original research paper	15	10th week of the course	Evaluated using 10 pt system. Criteria: understanding of the material, capability to transfer the knowledge to the fellow colleagues.
Midterm exam	30	9th week of the course	Multiple choice examination. 100 points total. The knowledge of the subject is evaluated as follows: >90 pts

			excellent, >85 points very good result, >75 pts good, >65 pts satisfactory, <50 pts unsufficient level.
Final Exam	30	16th week of the course	Multiple choice examination. 100 points total. The knowledge of the subject is evaluated as follows: >90 pts excellent, >85 points very good result, >75 pts good, >65 pts satisfactory, <50 pts unsufficient level.
Total	100		Weighted cumulative score.

Author	Year of publica-tion	Title	Issue of a periodical or volume of a publication	Publishing place and house or web link
Compulsory reading				
J. Pevsner	2009	Bioinformatics and Functional Genomics (2nd ed.)		Wiley-Liss
J. Gu, Ph. E. Bourne (eds.)	2009	Structural Bioinformatics (2nd ed.)		Wiley-Blackwell
Optional reading				
A. Lesk	Several editions	Introduction to Bioinformatics		Oxford University Press
Various authors		Lecture notes on the internet		For example, http://lectures.molgen.mpg.de/ online_lectures.html