

FACULTY OF CHEMISTRY

SUBJECT CARD

Name of subject in Polish **Zaawansowana Bioinformatyka**
Name of subject in English **Advanced bioinformatics**
Main field of study (if applicable): **Biosciences**
Specialization (if applicable): **Bioinformatics**
Profile: **academic**
Level and form of studies: **2nd level, full-time**
Kind of subject: **obligatory**
Subject code W03BSS-SM2010P
Group of courses NO

	Lecture	Classes	Laboratory	Project	Seminar
Number of hours of organized classes in University (ZZU)				30	
Number of hours of total student workload (CNPS)				60	
Form of crediting (Examination / crediting with grade)				crediting with grade	
For group of courses mark (X) final course					
Number of ECTS points				2	
including number of ECTS points for practical classes (P)				2	
including number of ECTS points corresponding to classes that require direct participation of lecturers and other academics (BU)				1	

*delete as not necessary

PREREQUISITES RELATING TO KNOWLEDGE, SKILLS AND OTHER COMPETENCES

1. Knowledge of biological basis related to bioinformatics
2. Proficiency in use of online bioinformatics databases of sequences, genes, genomes etc.
3. Proficiency with use of Unix systems, ability to work in the command line
4. Ability to solve algorithmic problems and automating computing tasks with use of a programming language like Python

SUBJECT OBJECTIVES

- C1 Gain of the skills necessary to use remote computing facilities ("cloud computing" services), including preparation and configuration of system images packaged with necessary software to be run on such services;
- C2 Gain of the skills necessary to document and archive the team work on the process of creation, testing and development of specialized software and workflows to process bioinformatical data sets, with use of interactive Jupyter notebooks, version control systems and online repositories;
- C3 Gain of the skills necessary to perform data processing, analysis and interpretation of the results of new generation sequencing experiments, including quality control and filtering (trimming), mapping on reference genomes, analysis and visualization;
- C4 Gain of the skill necessary to use GNU R package with selected Bioconductor modules to perform typical bioinformatics analyses and visualization of experimental results.

SUBJECT EDUCATIONAL EFFECTS

Relating to skills:

PEU_U01 Students can prepare and configure system images with software packages and services required to run bioinformatics related computation, data processing and analysis on remote computing facilities (“cloud computing”).

PEU_U02 Students can document and archive the performed analysis workflows and results and developed software codes using version control systems, online repositories and interactive notebooks.

PEU_U03 Students can perform full analysis of NGS sequencing results, from the initial data processing to mapping on the reference genomes, visualization and interpretation

PEU_U04 Students can use the GNU R system and Bioconductor libraries to perform typical analyses of experimental data sets available online, and to visualize the results.

Relating to social competences:

PEU_K01 Students can work in teams to solve problems and improve proposed solutions

PEU_K02 Students can communicate and describe the workflow and results of typical data analyses performed in bioinformatics.

PROGRAMME CONTENT

Project		Number of hours
Proj 1	Introductory classes: the form and organization of the classes, didactic materials, requirements for the final grade	2
Proj 2	Documentation and archiving of the performed analyses, workflows and developed software: <ul style="list-style-type: none"> • Version control system git and online repositories (github, gitlab etc.). Team working practices. • Documentation of the work using interactive Jupyter notebooks • Markdown syntax 	2
Proj 3	Project 1: Cloud computing <ul style="list-style-type: none"> • Preparation of system images for remote execution using Docker. Configuring network services and permanent storage. Simple servers of Jupyter, ssh, sshfs. • Preparation and configuration of Docker system image with all services needed to run the GALAXY environment. • Preparation and configuration of Docker system image with Jupyter or JupyterLab server, Jupyter R kernel, R system and selected Bioconductor modules. • Depositing of the results (Dockerfiles, notebooks) in a git repository. Working in teams to test solutions and solve problems. 	4
Proj 4	Project 2: Analysis of NGS sequencing data using the GALAXY platform <ul style="list-style-type: none"> • Different sequencing platforms and their relation to the results and their analysis • Characteristics of NGS data: read quality Q, FASTQ format variants, sequencing depth, filtering of low quality results. • Initial data processing (trimming). • Mapping results on the reference genome; analysis and visualization of results. 	10

Proj 5	Project 3: Analysis of experimental data sets using GNU R and Bioconductor <ul style="list-style-type: none"> • Introduction to use of the GNU R system: data types, plotting and visualization • Selected Bioconductor modules and functions • Characteristics of experimental data depending on the platform, required processing • Analysis of example data sets 	10
Proj 6	Students presentations of their projects	2
	Total hours	30
TEACHING TOOLS USED		
<p>N1. Instructions and video recording for self-study prior to the relevant classes (the “reverse classroom” approach).</p> <p>N2. Multimedia presentations and live demonstration how to use software.</p> <p>N3. Problem solving – individual and in teams – with the help of the tutor and using online resources.</p> <p>N4. Use of specialized software for performing typical analyses of bioinformatics data sets .</p> <p>N5. Students presentations of their work, solved problems and results.</p>		

EVALUATION OF SUBJECT LEARNING OUTCOMES ACHIEVEMENT

Evaluation (F – forming during semester), P – concluding (at semester end)	Learning outcomes code	Way of evaluating learning outcomes achievement												
F1	PEU_U01, PEU_U02, PEU_K01	Project 1 score												
F2	PEU_U03, PEU_K02	Project 2 score												
F3	PEU_U04, PEU_K02	Project 3 score												
P = F1 + F2 + F3		<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Score</th> <th style="text-align: left;">Grade</th> </tr> </thead> <tbody> <tr> <td>50-59,99%</td> <td>3,0</td> </tr> <tr> <td>60-69,99%</td> <td>3,5</td> </tr> <tr> <td>70-79,99%</td> <td>4,0</td> </tr> <tr> <td>80-89,99%</td> <td>4,5</td> </tr> <tr> <td>90-100%</td> <td>5,0</td> </tr> </tbody> </table>	Score	Grade	50-59,99%	3,0	60-69,99%	3,5	70-79,99%	4,0	80-89,99%	4,5	90-100%	5,0
Score	Grade													
50-59,99%	3,0													
60-69,99%	3,5													
70-79,99%	4,0													
80-89,99%	4,5													
90-100%	5,0													
PRIMARY AND SECONDARY LITERATURE														

PRIMARY LITERATURE:

Due to rapid progress of the relevant technologies, the best sources of information are the online learning resources and software documentation:

- [1] <https://git-scm.com/doc>
- [2] <https://docs.docker.com>
- [3] <https://docs.jupyter.org>
- [4] <https://training.galaxyproject.org>
- [5] <https://cran.r-project.org/doc/manuals>
- [6] <https://bioconductor.org/help>
- [7] Relevant articles on <https://wikipedia.org>

SECONDARY LITERATURE:

- [1] „Next-generation sequencing : current technologies and applications”, ed. Xu, Jianping; Caister Academic Press, Norfolk 2014. ISBN 978-1-908230-33-1
https://omnis-pwr.primo.exlibrisgroup.com/permalink/48OMNIS_TUR/d7ok8p/alma9960747679207668
(mostly of historical interest due to rapid progress of NGS technology and software, but it does introduce the basic background and concepts)
- [2] <https://socviz.co/gettingstarted.html> (Introduction to RMarkdown as a tool to document an analysis workflow and results using the R system)

SUBJECT SUPERVISOR (NAME AND SURNAME, E-MAIL ADDRESS)

Paweł Kędzierski, pawel.kedzierski@pwr.edu.pl