МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Біологічний факультет
Кафедра генетики та біотехнології

Затверджено
на засіданні кафедри генетики та біотехнології
(протокол № 17 від 29 серпня 2025 р.)

Завідувач кафедри._____
Віктор ФЕДОРЕНКО

Силабус з навчальної дисципліни «Інформаційні технології та бази даних у біотехнології» (англійською мовою), що викладається в межах ОПП «Біотехнології та біоінженерія» другого (магістерського) рівня вищої освіти для здобувачів зі спеціальності G21 Біотехнології та біоінженерія

Львів 2025

Course title	Information technologies and databases in biotechnology
Location	Hrushevskoho st 4, Lviv 79005
Faculty and Department	Biological Faculty, Dept. of Genetics and Biotechnology
Area, code, specialty	G Engineering, manufacturing and construction G21 Biotechnologies and bioengineering
Lecturer	Bohdan Ostash (lectures and practicals)
Contact info	bohdan.ostash@lnu.edu.ua 032 2394407 http://bioweb.lnu.edu.ua/employee/ostash-b-o
Consultations	*In person*: II semester (2025), every Tuesday, 11:30-13:00 *Online:* in the form "question-answer" via email, in working days of the week,10:00-16:00; please allow for a three days since the moment of emailing the question.
Course webpage	https://bioweb.lnu.edu.ua/course/itdbb
Course information	The course is developed so that MSc absorb professionally-oriented knowledge of available services and databases in the area of biomedical data that are not part of general course of bioinformatics. Following topics will be detailed – descriptive and inferential statistics, PubChem, Taxonomy, VAST+, ortholog databases, Galaxy, tools to perform principal component analysis and heatmap generation, illustration of statistical data. The course includes lectures and practicals where the student will be introduced to the available web tools and services.
Course annotation	As of 2024 the databases included the information about genes of over 400 000 organisms, which is over 21×10^{12} base pairs. This dataset is linked to its derivatives, such as information on expression levels, effector molecules, sructures of the encoded proteins, phylogenetic trees, epidemiological and population data. Systematic analysis of the entire dataset is possible through the use of computational methods, and judicious application of statistical assessment to the results. This course focuses on biotechnologically relevant datasets and statistical methods.
Goal and tasks of the course	**Goal:** to develop in students a system of knowledge about main databases and services which can be used for the analysis of various biomedical data. **Tasks:** *a*) to introduce to variety of biomedical databases and statistics; *b*) develop an active knowledge about scientific problems which can be addressed with these databases and tools.
Literature to learn the discipline	**Main literature:** **Primary:** 1. NCBI overview. Online access: https://www.nlm.nih.gov/portals/researchers.html 2. Statistics for biologists. https://www.nature.com/collections/qghhqm **Additional:** 3. Pevsner J. Bioinformatics and functional genomics. 3rd edition. Wiley Blackwell, UK. – 2015- 1116 p. ISBN 978-1-118-58178-0.
Course duration	One semester
Course content	Lectures – 16 h, practicals – 16 h, student work – 58 h; total – 90 h (3 ECTS credits)
Expected results of the learning	After this course MSc will **know**: - principles and approaches of descriptive statistics; - methods of inferential statistical analysis;

	- databases and services od biotechnological type. **and will be able to do**: - to use contemporary approaches and services of statistical assessment of the data; - to use databases where the information on nucleotide and amino acid sequences, chemical structure of the compounds is stored; - to identify regulatory sequences in genomes; - to optimize genetic sequences for the expression.
Common competencies, Professional competencies, Program results of learning	The course is developed so that it forms following common (CC) and professional competencies (PC): CC01. Ability to work in the international context. CC02. Ability to use information and communication technologies. CC03. Ability to generate new ideas (creativity). PC01. Ability to make use of novel findings of biology, necessary for professional, research and/or innovative activity. PC02. Ability to outline a modeling task, to generate process and object models at different levels of organization of living matter with the help of mathematical models and information technologies. PC03. Ability to use modern information technologies and to analyze the information in the area of biology and across the disciplines. PC06. Ability to predict the directions of development of contemporary biology on the basis of generalized analysis of the state of science and technologies. PC07. Ability to diagnose the state of biological systems on the basis of investigation of the organisms on different organization levels. PC10. Ability to use the results of research in practical activity. Program results (PR) of the course: PR1. To know state and foreign languages enough for fluent professional communication and presentation of research results. PR2. To use libraries, information databases, web resources to find necessary information. PR4. To solve complex biological problems, generate and evaluate the ideas. PR5. To analyze and evaluate the impact of achievements in the area of biology on the societal development. PR6. To analyze biological phenomena and processes on the molecular, cellular, organismal, population and biome levels from the point of view of fundamental scientific knowledge, and to use for this specialized modern research methods. PR11. To carry out statistical tests, analysis and summary of the experimental data using up-to-date software products and information technologies. PR14. To adhere to academic integrity norms during the learning and research activities; to know basic legal norms with regard to intellectual property protection. PR16. Critical assessment of theories, principles, methods from various branches of biology for solving practical tasks and problems.
Keywords	Mathematical models in biology, databases, statistical analysis.

Course Format	In-person or remote
Themes	See the table at the bottom of this syllabus
Final control, form	Exam in the of the semester (oral).
Prerequisites	Basic level of English; knowledge of principles of genetics, biochemistry, zoology and botany. Acquaintance with basic math concepts such as log, odds, exponent, per cent values; basic knowledge of probability theory and statistics. User level of handling the laptops. Internet connection.
Teaching methods to be used in this course	Lectures, explanations, work on control assignments during practical lessons.
Equipment	Laptop or tablet, common computer programs (internet browser, preferably Mozilla Firefox or Google Chrome; Excel, some text editor)
Assessment criteria	Score during the semester/Exam scores – 50/50 Content of the first five lectures will the subject of the semester control work. The latter will include two questions: (graded max. 5 points each). Maximal score for the control work is 10 points. Writing the control work is obligatory. Coursera certificate in biostatistics will cover the required control work, with max score: https://www.coursera.org/learn/biostatistics. Equivalent courses will be considered on case-by-case basis – please feel free to ask! A student can get up to 40 points by carrying out eight assignments during the practicals (DABEST, PCA+Heatmap, ExPaSy, Galaxy, PubChem, UCSF Browser, CR-BRV, DTU services). Admission to the exam requires at least 25 points, of which no less than 5 were for the semester control work. The exam includes questions, term definitions and tests covering the entire lecture and practical course. A student will have 30 min to prepare for exam, then answers are presented to the lecturer orally. Zero tolerance to any form of violation of academic integrity, as laid out in https://surl.li/vtajmp
Exam questions	1. Types of data for statistical analysis 2. Differences between qualitative (nominal and ordinal ones) and quantitative (continuous and discrete) data 3. What can you measure in your data? 4. What are data' central tendency and their spread? 5. Why does uncertainty assessment matter? 6. Population, generalized sample, sample and the Gauss theorem 7. Sampling process in statistic analysis. Sample size 8. Random vs non-random sampling 9. Confidence intervals and distributions. 10. Statistical inference: null hypothesis 11. The test statistic and P value 12. Type I and type II errors. Power 13. Examples of statistical tests 14. T test 15. Chi-squared test 16. Pearson correlation coefficient 17. Spearman correlation coefficient 18. PubChem 19. Galaxy

	20. PCA tools 21. Heatmap tools 22. Epidemiological tools based on NGS data 23. DTU server 24. Taxonomy databases 25. Homology finding severs
Final survey	Will be offered in the end of the course

Table

Scheme of the course. Mode of learning: full attendance

Week	Theme	Activities, hours	Additional resources	Duration
1,2	Introduction. Types of biotechnological data	Lectures – 2 h. Practicals – 2 h. Student work – 7 h.	https://www.estimationstats.com/#/background	2 weeks
3,4	Descriptive statistic. Sample and its size	Lectures – 2 h. Practicals – 2 h. Student work – 7 h.		2 weeks
5,6	Descriptive statistics. Confidence intervals	Lectures – 2 h. Practicals – 2 h. Student work – 8 h.		2 weeks
7,8	Inferential statistics. Statistical tests	Lectures – 2 h. Practicals – 2 h. Student work – 8 h.		2 weeks
9, 10	Databases PubChem, Taxonomy	Lectures – 2 h. Practicals – 2 h. Student work – 7 h.	https://pubchem.ncbi.nlm.nih.gov/	2 weeks
11, 12	Galaxy – biomedical data analysis	Lectures – 2 h. Practicals – 2 h. Student work – 7 h.	https://galaxyproject.org/support/	2 weeks
13, 14	Principal component analysis; heatmaps	Lectures – 2 h. Practicals – 2 h. Student work – 7 h.	https://www.chiplot.online/ https://biit.cs.ut.ee/clustvis/	2 weeks
15, 16	Genetic data analysis for epidemiology purposes. Homology finders	Lectures – 2 h. Practicals – 2 h. Student work – 7 h.	https://www.genomicepidemiology.org/ https://link.springer.com/article/10.1007/s10709-023-00196-8	2 weeks

Author Bohdan OSTASH

"Approved"

Head of teaching curricula commission
of biological faculty
Vitaliy Honcharenko

"29th"August 2025

Guarantor of EPP «Biotechnologies and bioengineering»
Bohdan OSTASH

29 August 2025 p.