<table>
<tr><td colspan="2">NAME OF THE COURSE</td><td colspan="5">Bioinformatics</td></tr>
<tr><td>Code</td><td>PMB704</td><td>Year of study</td><td colspan="4">1.</td></tr>
<tr><td>Course teacher</td><td>Željka Trumbić, PhD, Assistant Professor
Matilda Šprung, PhD, Assistant Professor</td><td>Credits (ECTS)</td><td colspan="4">5</td></tr>
<tr><td rowspan="2">Associate teachers</td><td rowspan="2"></td><td rowspan="2">Type of instruction (number of hours)</td><td>L</td><td>S</td><td>E</td><td>F</td></tr>
<tr><td>15</td><td></td><td>45</td><td></td></tr>
<tr><td>Status of the course</td><td>Compulsory</td><td>Percentage of application of e-learning</td><td colspan="4"></td></tr>
<tr><td colspan="7">COURSE DESCRIPTION</td></tr>
<tr><td>Course objectives</td><td colspan="6">Bioinformatics represents a broad interdisciplinary field formed at the interface of revolutionary development of molecular biology and computer science at the beginning of the 21st century, as a result of the need to interpret vast amount of data coming from large scale molecular biology efforts, such as genomics, proteomics, transcriptomics, etc. The focus of bioinformatics is the use of computer databases and algorithms to analyze individual genes, or the entire DNA makeup of an organism, the genome, individual proteins or complex functional networks of macromolecules and biochemical pathways. The goal of this course is to familiarize the students with some of the basic tools and databases in bioinformatics, specifically associated to transcriptomics and protein analyses, providing solid bases for students to develop skills needed for independent experiment design, data collection and analyses, as well as independent pursue of initiatives to expand their knowledge in this broad and dynamic field.</td></tr>
<tr><td>Course enrolment requirements and entry competences required for the course</td><td colspan="6">Fundamental background in molecular biology and genetics.</td></tr>
<tr><td>Learning outcomes expected at the level of the course (4 to 10 learning outcomes)</td><td colspan="6">After completing the course, the students should be able to clearly:
• State main biological databases and know how to search them
• Define what BLAST algorithm is, compare a sequence to a database using BLAST algorithm and interpret the results
• Define the transcriptome and understand the methodology used for its study (RNAseq)
• Design a good transcriptomic profiling experiment
• Use some of the basic bioinformatic tools (UNIX command line, R packages, RStudio)
• Analyze RNAseq experiment and interpret the results of differential gene expression analyses
• List and name the basic elements of protein structure
• Perform multiple sequence alignments of proteins and design protein model based on homolog structure
• Visualize and analyze proteins using bioinformatics tools such as PyMOL, VMD or similar programs for macromolecular visualization</td></tr>
<tr><td>Course content broken down in</td><td colspan="6">Lecture 1. Introduction to bioinformatics and biological databases. (2 hours)</td></tr>
</table>

detail by weekly class schedule (syllabus)	Definition of bioinformatics and domains of –*omics* research: genomics, transcriptomics, proteomics, etc. Overview of databases at the NCBI (National Center for Biotechnology Information) portal, EBI (European Bioinformatics Institute) and access to sequences. Lecture 2. FASTA format for storing sequences. BLAST algorithm (2 hours) Brief recap of Sanger sequencing, storing sequences in FASTA format. Definition and description of the BLAST algorithm. Lecture 3. Transcriptome definition, description and regulation (2 hours) Biology of transcription. Introns, exons, alternative splicing and isoforms. Transcriptome composition: coding and non-coding RNA in mutual interaction that dynamically shapes the transcriptome. Lecture 4. Transcriptome profiling by sequencing – RNAseq (2 hours) Classical methods of transcriptome composition and gene expression analyses vs sequencing: pros and cons. Different types of RNAseq libraries (*stranded* vs *unstranded*, *3' tagged RNA-seq).* Long and short read sequencing, paired-end vs single-end sequencing. Lecture 5. How to design a good experiment for transcriptome profiling? (1 hour) Factors to consider when designing experiments, preparing samples for the analyses and sequencing that might be the source of batch effects hampering data analyses. RNAseq workflow steps: from experiment design and set-up to data analyses. Lecture 6. (2 hours) Biology of translation and protein structure. Translation of genetic information at ribosomes. Proteins and amino acids. Peptide bonds and other important chemical interactions in proteins. Protein structure and folding. Lecture 7. (2 hours) Pairwise and multiple alignment. Protein homology. Design of protein models based on homology protein structure. Identification of conserved residues and analysis of their function. Lecture 8. (2 hours) Protein databases and modelling of macromolecules. Searching ExPASy and PDB. Withdrawal of protein sequences and visualization of proteins in PyMOL or VMD. Exercise 1. Database searching. FASTA format and sequence manipulation in program MEGA. Comparing sequences to databases using BLAST algorithm. BLAST for multiple sequences. (3 hours) Exercise 2. Introduction to UNIX (6 hours)

	Brief introduction to the UNIX operating system and the command line. Navigating through the system (commands pwd, ls, cd), creating, copying, deleting files and directories (mkdir, cp, rm), searching files (find), displaying files (head, tail, cat, less), pattern searching (grep). Running processes. Exercise 3. Introduction to FASTQ files containing sequencing results. Using UNIX commands to manipulate FASTQ files. (3 hours) Exercise 4. Quality control of sequencing reads and quality trimming (3 hours) FastQC program will be used to check for sequencing run quality and distinguish a poor from a good run. Students will subsequently perform quality trimming of the reads in order to remove adapter sequences and bases that fall below a minimum quality threshold. Exercise 5. Aligning reads to a reference genome. Alignment results in SAM/BAM formats and visualization of the alignment using IGV (Integrative Genomics Viewer). Defining gene models and counting reads (3 hours) Exercise 6. Introduction to R. (6 hours) Introduction to R, a free software environment for statistical computing and graphics. Introduction to RStudio interface with R. Creating projects in RStudio and data organization. Basic data objects in R (vectors, matrices, data frames, lists). Reading data into R, data cleaning and manipulation (package dplyr). Bioconductor packages for the analysis and comprehension of high-throughput genomic data. Exercise 7. Differential gene expression (6 hours) Exploratory data analyses using PCA (principal component analyses) in R. RNAseq data normalization in respect to library size and gene size. Statistical models used for testing differential expression based on negative binomial distribution (packages EdgeR and DESeq2). Multiple testing corrections (FDR) and visualization of differentially expressed genes. Exercise 8. Gene set enrichment analyses: KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology) sets. Visualization of results using tables and graphs (3 hours) Exercise 9. What to do if the referent genome isn't available yet? (2 hours) Brief introduction to de novo transcriptome assembly (program Trinity), searching for open reading frames and coding regions. Transcript annotation using BLAST. Programs for read alignment to the reference transcriptome (RSEM, Salmon) Exercise 10. Pairwise and multiple alignment. Design of protein homologs using multiple alignments and identification of candidate sequence. Using ExPASy tools develop a homology model. Translate model in interface for molecular

<table>
<tr><td></td><td colspan="6">visualization. Analyzing protein structure and identification of residues at important positions in the protein sequence. (3 hours)

Exercise 11.
Protein databases. ExPASy and PDB. Searching for relevant proteins and obtaining all available data. (3 hours)

Exercise 12. Modelling of macromolecules
Using ExPASy tools build macromolecular models and analyze their structures. Identification of most important residues and their bonding in the structure. Position of active site and residues important for catalysis. (4 hours)</td></tr>
<tr><td>Format of instruction</td><td colspan="3">☒ lectures
☐ seminars and workshops
☒ exercises
☐ on line in entirety
☐ partial e-learning
☐ field work</td><td colspan="3">☒ independent assignments
☒ multimedia
☐ laboratory
☒ work with mentor
☐ (other)</td></tr>
<tr><td>Student responsibilities</td><td colspan="6">The course is organized into lectures and practical exercises (PowerPoint presentations, educational video clips, practical computer training). It is expected that students will attend the classes regularly and the attendance will be recorded. Students are expected to actively participate during classes, successfully conduct the final project and pass the oral exam. Should students be prevented from meeting course requirements due to unexpected circumstances, they should notify the teachers in a timely fashion.</td></tr>
<tr><td rowspan="6">Screening student work (name the proportion of ECTS credits for each activity so that the total number of ECTS credits is equal to the ECTS value of the course)</td><td>Class attendance</td><td>1</td><td>Research</td><td></td><td>Practical training</td><td>2</td></tr>
<tr><td>Experimental work</td><td></td><td>Report</td><td></td><td>(Other)</td><td></td></tr>
<tr><td>Essay</td><td></td><td>Seminar essay</td><td></td><td>(Other)</td><td></td></tr>
<tr><td>Tests</td><td></td><td>Oral exam</td><td>1</td><td>(Other)</td><td></td></tr>
<tr><td>Written exam</td><td></td><td>Project</td><td>1</td><td>(Other)</td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Grading and evaluating student work in class and at the final exam</td><td colspan="6">Student attendance will be recorded during each class. Students will be encouraged to actively contribute to the course and independently solve problems, allowing them to score points towards the final grade. At the end of the course, students will choose a project of their own interest in transcriptomics or protein analyses in order to demonstrate independent ability to use bioinformatics tools in practice. Students will finalize the project in consultation with the teachers and present final results during the oral exam. Final grade will be allocated based on total student effort during the course, active participation, project performance and the oral exam discussion. Evaluation and final grade:
90% - 100%: grade 5 (excellent);
80% - 89%: grade 4 (very good);
65% - 79%: grade 3 (good);
51% - 64%: grade 2 (sufficient);
< 51%: grade 1 (insufficient).</td></tr>
</table>

<table>
<tr><td rowspan="4">Required literature (available in the library and via other media)</td><th>Title</th><th>Number of copies in the library</th><th>Availability via other media</th></tr>
<tr><td>Bioinformatics and Functional Genomics, third edition. 2015. Jonathan Pevsner. Wiley-Blackwell.</td><td>1</td><td></td></tr>
<tr><td>Introduction to Bioinformatics, 3rd Edition. 2008. Arthur M. Lesk. Oxford University Press.</td><td>1</td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>Optional literature (at the time of submission of study programme proposal)</td><td colspan="3">1. Love MI, Anders S, Kim V and Huber W. RNA-Seq workflow: gene-level exploratory analysis and differential expression [version 2; peer review: 2 approved]. F1000Research 2016, 4:1070 (https://doi.org/10.12688/f1000research.7035.2)
2. Zuur AF, Ieno EN, Meesters EHWG. A beginner’s guide to R. 2009. Springer, New York.
3. Trumbić Ž. 2019. Internal protocols – practical courses in Bioinformatics. University of Split.
4. Trumbić Ž. 2019. Additional teaching materials – powerpoint presentations and supporting material for practical courses.</td></tr>
<tr><td>Quality assurance methods that ensure the acquisition of exit competences</td><td colspan="3">Lectures and exercises are conceived as interactive activities in the classroom where active participation of students will be encouraged through questions, comments and independent task solving, enabling continuous monitoring of students' work and understanding. Analysis of the level of acquisition of exit competences will be performed through final project and oral exam.</td></tr>
<tr><td>Other (as the proposer wishes to add)</td><td colspan="3">Consultation hours and exam dates will be publically available through the website of Faculty of Science Split. Students are free to contact teachers by e-mail at msprung@pmfst.hr and ztrumbic@unist.hr to discuss any questions or need additional support in meeting course requirements.</td></tr>
</table>