## *Course memo – Kurs-PM*
## Bioinformatics CB2442 (7.5 hp)

### Canvas
Canvas will be used for communication during the course, including updates to this course PM. Your access to the **CB2442** event at **https://canvas.kth.se** will be activated after course registration. Log in using your kth.se-account and look for the event **CB2442 Bioinformatics** for the semester you take the course. Course material and information, e.g., lecture notes and lab instructions, will be provided through Canvas.

### Aim
The aim of the course is that the students should acquire sufficient knowledge about fundamental bioinformatics and programming resources and methods, so as to know how to (i) explain the theory behind the methods, (ii) use the corresponding tools and resources to perform analyses of molecular biology data, and (iii) interpret the results of these analyses.

Specifically, after passing the course, the student should be able to:
- explain basic bioinformatic methods
- give an account of applications and limitations of bioinformatic methods
- use relevant methods to analyze bioinformatic problems
- interpret the results of bioinformatic analyses
- program basic bioinformatic algorithms

### Syllabus
The course is organized in 4 modules:
1. Alignment [5 lectures]
2. Sequence features [7 lectures]
3. Phylogenetics [3 lectures]
4. Computer exercises [8 exercises: 4 Python programming, 4 bioinformatics tools usage]

There is also an Introductory lecture (combined with the first Sequence feature lecture) and a Q&A session. In total 16 lectures and 8 computer exercises.

The course credits are distributed into 2 parts:
- Examination (TEN1, grade scale A-F), 5 hp
- Computer exercises (LAB1, grade scale P/F), 2.5 hp

### Course contents
Bioinformatics: Pairwise sequence alignment of protein and DNA/RNA sequences, multiple sequence alignment, significance of alignment results, features of protein and DNA/RNA sequences including sequence conservation and homology, gene expression, differential gene expression, clustering of vectors, phylogenetic analysis, introduction to relevant public databases and extraction of relevant data from these. How to use command line operations for data analysis and file management (in Unix/Linux, or corresponding operating systems), as well as available web-based tools for data analysis.
Programming: Programming of basic bioinformatics algorithms in Python, using basic operations on data such as conditional execution, loops, and regular expressions.

The course content is the union of all material presented at the lectures, at the computer exercises, and in the course literature as specified by the reading assignments, as well as any additional handouts or other resources specified during the course.

### Teachers
Olof Emanuelsson, Lukas Käll, Anders Andersson

### Teaching assistants on computer labs (TAs):
See Canvas

### Prerequisites (recommended)

BB1150 Biochemistry 1 (or equivalent), BB1190 Gene technology (or equivalent), SF1625 Calculus in one variable (or equivalent), SF1911 Statistics for bioengineering (or equivalent), and BB1000 Programming in Python (or equivalent).

**Literature**

*Web-based bioinformatics textbooks:*

Bioinformatics for Biotechnology students. Lukas Käll & Anders Andersson
Available online at: https://www.kaell.se/bibook

Computational Biology - Genomes, Networks, and Evolution. Manolis Kellis, et al.
Available online and as a (free) downloadable pdf at:
https://bio.libretexts.org/Bookshelves/Computational_Biology/Book%3A_Computational_Biology_-_Genomes_Networks_and_Evolution_(Kellis_et_al.) (or search for "libretexts kellis", this will be the first hit)

*Python resources (non-mandatory) for programming labs:*

Online resource: Jake VanderPlas: Whirlwind Tour of Python
http://nbviewer.jupyter.org/github/jakevdp/WhirlwindTourOfPython/blob/master/Index.ipynb

Book: Al Sweigart: Automate the Boring Stuff with Python Programming.

VS Code introduction video: Learn Visual Studio Code in 7 min.
https://code.visualstudio.com/docs/introvideos/basics

*Articles:*

Eddy: Where did the BLOSUM62 alignment score matrix come from? *Nature Biotechnology*, vol **22**, pp 1035-1036 (2004)

Deshpande et al.: RNA-seq data science: From raw data to effective interpretation. *Frontiers Genet* 14:997383, 2023: pages 01-07, left column only of page 07

Pop: Genome assembly reborn. Briefings in bioinformatics 10:354-366, 2009, excluding "Eulerian path", "Information integration", "Metagenomics/mixtures of organisms".

Sherlock: Analysis of large-scale gene expression data. *Current Opinion in Immunology*, vol **12**, pp 201-205 (2000)

Holder & Lewis: Phylogeny Estimation: Traditional and Bayesian Approaches. *Nat Rev Genetics,* vol **4**, pp 275–284 (2003) (the part about Bayesian phylogenetics is optional)

Additional articles or other resources may be specified during the course.

Detailed reading assignments (chapters in books, sections in articles, additional resources etc) will be provided during the course. Handouts, lab instructions, articles, and other resources specified during the course are also part of the required reading.

**Requirements for examination**

A. **Written exam (TEN1).** 5.0 hp. 4 hours written exam with theory questions and problem solving. Grades: A, B, C, D, E, Fx, F, where A-E correspond to pass, F is fail, and Fx is fail with the possibility to complement.
B. **Computer exercises (LAB1)**. 2.5 hp. Grades: Pass/Fail. Pass means that all computer exercises have been performed and that a report for each exercise has been handed in *in time* and has been approved.

To receive a final grade at the course you must pass both the written exam (with grades A-E) and receive a Pass grade at the computer exercises. Your final grade will be the same as the grade on the written exam. If you receive an Fx at the written exam you may request to be assigned a complementary assignment to complete and hand in within the stipulated time, and if deemed satisfactory your grade will be changed from Fx to E. If you receive a Fail at the computer exercises, you must consult with the teaching assistants to find out what you need to do to change that into a Pass.

**Active participation in Quiz sessions may give you bonus points on the exam**

For some of the lectures in the three bioinformatics modules (Alignment, Sequence features,

Phylogenetics) you are expected to prepare in advance by reading the provided material and/or watching the provided video recordings. At these lectures there will be a Quiz session (QS), with questions on the provided material. First, you reply individually to the first question in the quiz. Second, you discuss the question and your answer with your peers (in pairs) in the lecture room. Third, the lecturer will ask a couple of students in the lecture room (based on the names of students signed in to the quiz) to explain for the class how they approached and solved the question. These steps are repeated for each question in the quiz.
Participation (including preparation, active discussion with your peers and explanation of solution [if selected]) at these Quiz sessions will give you bonus points on the exam. Note that you will have to run the quiz to receive bonus points, and answer all questions in the quiz. Instructions on how to access the quiz will be given at each session. You will need a smart phone with QR code reading capability.
There will be in total **9 Quiz sessions** at 9 different lectures:

- Alignment module: **4** Quiz sessions.
- Sequence features module: **3** Quiz sessions.
- Phylogenetics module: **2** Quiz sessions.

Participation in **7-9** of the Quiz sessions will give you **4 bonus points.** Participation in **5-6** will give **2 bonus points.** (Participation in <5 will not give any bonus points).

### The final written exam

The final written exam will contain theory questions and problem solving based on the contents of the course. You are allowed to use a calculator, including graph calculator, where the memory is empty at the start of the exam. The exam consists of two parts, A and B. Part A consists of 12 multiple-choice questions (one alternative per question) worth 2 points per question (either 0 or 2). Part B consists of 4 essay questions worth 5 points per question. To pass the exam (grade E) you need at least 16 points on Part A (8 correctly answered questions). **If you have less than 16 points on Part A, the Part B questions will not be corrected**. The final grade (given that you pass Part A) will be based on all your points (Part A + Part B). Bonus points from the Quiz sessions will be counted into Part A (and into the total score if you pass Part A). Thus, if you received bonus points from the Quiz sessions, you need at least 6 or 7 correctly answered questions on Part A in order to pass. If you did not receive bonus points, you need at least 8 correctly answered questions on Part A.
*NOTE 1*: **You must sign up for the final written exam in advance.**
*NOTE 2*: Rules in order to be allowed into the room where the exam is given:

- A valid **photo identification** is required, no exceptions.
- Be present 10 minutes before actual exam starting time: Entry to the exam is organized in two intakes. Intake 1: **10 minutes *before* the start**. Intake 2: 30 minutes after start. No admission in between, or after intake 2.
- If you haven't registered for the exam you may still be allowed to take it provided: (i) there is an available seat; (ii) you present a valid course registration certificate upon admission; (iii) you adhere to all other rules including the two mentioned above.

### Teaching activities

The **lectures** will consist of teacher-led lecturing of concepts as well as various student activities, including, for certain lectures, the Quiz sessions.

The **computer exercises** should be carried out in teams of two students and conducted **on site** in Linux/Ubuntu rooms at KTH Campus. You will be responsible for finding your own lab partner, and can use the Discussion session “Lab partner wanted” on Canvas for this.

The Programming part (LAB_P1 - LAB_P4) will be conducted using the VS Code environment. If you are not familiar with VScode, watch the video introduction (see Literature) *before* the first programming lab. When your team is finished with the Python programming task (when your script seems to work as it should) you should find another team that is also finished, sit together with this team and show and explain your respective codes to each other. The idea here is that you should see that programming tasks can be solved in different ways, and get ideas for how to solve other problems. When both groups have demonstrated their code (and potential remaining bugs have been fixed), you should contact a TA and go through your team’s code with the TA. If the TA

finds your code working as it should, you will be asked to submit your code to the Canvas report submission system and will get a Pass on the exercise, unless the TA after double checking the code on Canvas finds something that needs correction. Due to the setup of these exercises, it is important that you participate **on-site**. If you for some reason cannot participate on site, contact the course responsible. In this case you will be asked to conduct the lab remotely and come to another lab occasion to get your code assessed by a TA.

LAB_P1: Introduction to Python programming in VScode
LAB_P2: Programming multiple sequencing alignment
LAB_P3: Programming gene expression analysis
LAB_P4: Programming a phylogenetic tree / clustering algorithm

The exercises in the Bioinformatics part (LAB_B1 - LAB_B4) are also carried out in teams of two (may be the same teams as for the programming labs) and **on-site** in the Linux/Ubuntu rooms. For the LAB_B1 - LAB_B4 labs, you should hand in a written lab report *within two weeks after the exercise* using the Canvas report submission system. (Thus, you do not need to discuss with another team nor report your work on site to a TA).

The 4 bioinformatics exercises all revolve around a common scenario (an infection outbreak at a hospital), which will be available in Canvas. You must read this scenario before the first bioinformatics exercise. The bioinformatics labs will start with discussing a few **preparatory questions.** You will not gain or lose points from them, but you might be called to discuss one of them in front of your classmates, so be prepared. The preparatory questions are given in the beginning of each bioinformatics exercise instruction.

LAB_B1: Bioinformatics. Gene finding, Blast, and sequence alignment
LAB_B2: Bioinformatics. rRNA finding, taxonomic classification, and multiple sequence alignment
LAB_B3: Bioinformatics. Protein sequence feature prediction, multiple sequence alignment, and phylogenetics
LAB_B4: Bioinformatics. RNA sequencing, differential gene expression

Exercise instructions and data, for both bioinformatics and programming labs, will be provided through the Computer exercises module on Canvas.

### On site or online teaching?

**Lectures will be on site**. However, most lectures are pre-recorded and made available through Canvas in advance of the scheduled lecture time. In these cases, the scheduled lecture time will be used to discuss the contents of the recorded lecture and/or to do exercises related to the lecture contents. At some scheduled lecture time there will be Quiz sessions based on the provided material, that can generate bonus points on the exam (see above).

**Computer labs will be on site in Linux/Ubuntu rooms at KTH Campus.** We suggest that you use the Linux computers in the rooms. If you prefer to use your own laptop computer, you are responsible for installing the required software. Since it will take some time to set up your computer for the lab work, we require that you do this *in advance of the first lab*. See the "Software installation instructions" in the Computer exercises module in Canvas if you plan to use your own computer.

### Use of Generative AI

All use of generative AI is prohibited in the course's examination and graded assignments. This is based on the fact that the use of generative AI may have a negative impact on students' ability to meet the course's learning objectives.

The ban on generative AI means that

- All examinations, computer exercises, and graded assignments must be carried out without the help of generative AI.
- All submissions must be completely human-generated.
- This also includes that all debugging in code should be performed without generative AI.

As a student, you are fully responsible for all material you submit and must be able to defend and explain it completely without the support of generative AI.

**Disciplinary action for unauthorized use of AI**
Everything you submit must be your own work. Using generative AI in the course's examination and graded assignments is considered unauthorized assistance and may result in disciplinary action.

**Contact and information**
Initial information regarding the course is provided in this Course PM. All subsequent communication, including updates to the Course PM, will be communicated through the **CB2442 Bioinformatics** event in Canvas (**https://canvas.kth.se).**

See https://www.kth.se/student/kurser/kurs/CB2442?l=en for contact information.

**Disability**
If you have a disability, you may receive support from Funka:
https://www.kth.se/en/student/stod/studier/funktionsnedsattning/funka
Funka does not automatically inform the teachers (not even the course coordinator). Therefore, we recommend that you also inform the teachers regarding any needs you may have.

**Schedule**
The schedule is available at https://kth-gt.github.io/cb2442/general/schedule, where you can find the most detailed and up-to-date version. The schedule is also available at https://www.kth.se/en/student/studier/schema/sok-schema-1.2214.