BC. CROSS-BORDER BIOINFORMATICS STATE EXAM TOPICS

Bioinformatics expertise

Bioinformatics basics and pairwise sequence alignment

- Provide an overview of the most important types of data that occur in bioinformatics and where they come from, i.e. how they are measured or created.
- What is the general aim and approach of pairwise sequence alignment? Also comment on the biological plausibility of alignments and how this is reflected in the scoring schemes that are usually applied.
- Explain the general idea of the Needleman-Wunsch algorithm and how it is implemented.
- What is the essential difference between global and local sequence alignment? What are the resulting differences between the Needleman-Wunsch and the Smith-Waterman algorithm?
- Explain the general task of searching sequence queries in large biological basics. What are the most common methods? Why is the Smith-Waterman algorithm not feasible for this task?
- Explain the basic workflow of BLAST.

Multiple sequence alignment and phylogeny

- What are multiple sequence alignments useful for? What can we see and compute from them?
- What is the computational complexity of dynamic programming-based alignment methods like Needleman-Wunsch and Smith-Waterman? Why are similar approaches not suitable for multiple sequence alignment or the alignment of queries to large sequence databases? Which methods are used instead for these two tasks?
- Provide an overview of methods for scoring multiple alignments.
- Provide an overview of methods for constructing phylogenetic trees. Explain in more detail what the main difference between UPGMA and neighbor joining is.
- Explain the basic idea and realization of maximum parsimony. Explain the concept of tree length and two methods for tree search.

Microarrays

- How does a microarray machine work? How do we get from mRNA concentrations in the cell to CEL files?
- What can you do to assess chip quality from a CEL file?
- What is a Volcano plot? How should such a plot look like? What information can you derive from it?

- The four steps from CEL files to gene expression. Sketch one or two methods for each of the steps.
- How do you proceed from a matrix of gene expressions (genes x samples) to a list of differentially expressed genes?
- What are typical the next steps in a microarray project after having the list of differentially expressed genes?

Structural bioinformatics

- Given a set of short reads (a few hundred bases each) from 50 copies of a bacterial genome. Sketch the three major “paradigms” for assembling the genome.
- What does “RNA-seq” mean? Sketch how it works. To which technology does it compete? Which one is better?
- In intrinsic gene finding, what are typical signals that a certain stretch of DNA is really a gene?
- How can microarrays be useful in finding SNPs and CNVs?
- How does threading work (in the context of protein 3D structure prediction)?

Biology expertise

Genomics

- Principles of DNA sequencing and its application in genomics
- Sequence databases and the common DNA/AA sequence formats
- Evolution of prokaryotic and eukaryotic genomes
- Genome/transcriptome assembly
- Annotation of prokaryotic and eukaryotic genomes

Diversity of Life

- Give us a BRIEF outline of phylogenetic system of living things, as it is gradually being revealed in current decades, much owing to molecular and phylogenetic methods. Pay particular attention to a) mutual positions of Archaebacteria, Eubacteria and Eucaryota; b) system of five eucaryotic realms; c) position of vascular plants, fungi and metazoan animals within the system of five realms.
- Brief summary of modern phylogenetic methods used for classifying living things. Why was phylogenetic methodology "revolutionary" compared to earlier approaches? How and why is the methodology useful in biogeography, ecology and conservation?
- Discuss the "dynamic" (or "disturbance dynamics) paradigm, currently prevailing in ecology and conservation. Could you speculate on scientific issues in ecology, which might be resolved using bioinformatic methods?
- Discuss the environs of the two cities where you study - České Budějovice and Linz - from the perspectives od biogeography and biodiversity conservation. In which

phytogeograohy / zoogeography realms and provinces are the cities located; what biomes and habitats can be found there; which important nature conservation localities are located nearby; why are these localities important regionally, continentally and globally.

Molecular Biology & Genetics

- DNA structure, DNA packaging, DNA replication and the genetic code.
- Transcription (prokaryotic vs. eukaryotic mRNAs) – eukaryotic mRNA processing.
- Translation (genetic code & tRNAs), ribosome cycle.
- Mechanisms of gene expression control (e.g. prokaryotic gene operons, localised mRNAs, transcription factors, epigenetics/ imprinting/DNA methylation/post-translational histone modifications, insulators, RNA editing, riboswitches, RNAi)
- What is recombinant DNA? Cloning and techniques involved (e.g. plasmids, cDNA, PCR, restriction enzymes, DNA sequencing etc.)

Informatics Expertise

Data Structures & Algorithms

- Algorithms: Time and Space complexity, example of algorithms and their complexity
- Basic Data Structures: (array, linked-list, stack, queue), implementations and usage
- Sorting algorithms (selection sort, insertion sort, bubble sport, quicksort
- Neural Networks basics (neuron, synapse, axon, dendrite, weight, threshold, activation (output) function
- Model of Neural Networks : Perception, typical topologies (Supervised vs Unsupervised), deep learning
- Neural Networks learning: a principle, generalization, over fitting (overtraining), training set, testing set and validation set, back propagation.
- Genetic algorithms: a principle and basic terms (individual, population, crossover, mutation, selection, fitness).

Bash Programming

- Regular expressions, what is it, why and how to use them, write real some examples.
- How to connect inputs and outputs of several commands in bash? Present examples and their common usage.
- grep, sed, awk - why we are using them? Present examples and their common usage.

Parallel Programming

- MPI parallelization: communication of threads, memory, execution (advantages and disadvantages)

- OpenMP parallelization: communication of threads, memory, execution (advantages and disadvantages)
- Strategies in parallel programming: input and output operations, splitting of the work load (especially in cycles), efficiency (timing, number of threads)
- Pointers and references, relations between arrays and pointers, pointer arithmetic.