

Program of Studies:	Master Program Bioinformatics
Name of the module:	Special Lecture Bioinformatics: Algorithms for Sequence Analysis
Abbreviation:	BI-BM-1
Subtitle:	-
Modules:	Lecture and tutorial
Semester:	2nd semester master / every summer semester
Responsible lecturer:	Prof. Dr. Sven Rahmann
Lecturer:	Prof. Dr. Sven Rahmann
Language:	English
Level of the unit/ Mandatory or not:	Graduate course / mandatory elective
Course type/weekly hours:	Lecture: 4 h (2 x 2h weekly) Tutorial: 2 h (weekly)
Total workload:	300 h = 96 h of classes and 204 h of assignments and private study, including exam preparations
Credits:	9
Entrance requirements:	<ul style="list-style-type: none"> - Basic knowledge of sequence analysis, as taught in the Bachelor's courses (e.g., basic dynamic programming algorithms for pairwise alignment) - Basic probability theory and distributions (Binomial, Poisson, Geometric, Exponential, Normal, etc.) - Solid programming skills (ideally Python, but any language will do; there will be several programming exercises during the course).
Aims/Competences to be developed:	<p>This course covers a subject that is relevant for computer scientists in general as well as in computational biology and bioinformatics: the analysis of (biological) sequences (or of bit sequences, strings, ...). Because every information stored in computers is eventually serialized as a bit sequence, the methods taught in this course are quite universal, even though emphasis is placed on biological applications.</p> <p>The course will convey the ability, given a sequence or set of sequences, to choose an appropriate framework and a concrete algorithm and its parameters to analyze the sequence(s) and answer questions about it.</p> <p>The students learn how to transfer algorithmic principles to concrete practical problems (transfer competence). The course also practices communication competence during the presentation of solutions of exercises.</p>

<p>Content:</p>	<ul style="list-style-type: none"> - exact pattern matching algorithms (simple patterns, no index) and their analysis, including algorithms based on automata and bit-parallel algorithms, - full-text index data structures (suffix tree, suffix arrays) and applications to bioinformatics, e.g. repeat discovery; succinct index data structures (FM index) and algorithms (backward search), - approximate pattern matching (simple patterns, with and without index) and read mapping as an application, - pairwise sequence alignment: algorithms, variations, speed-ups; derivation of score matrices from Markov processes; statistical evaluation of alignments, - multiple sequence alignment: models, hardness, heuristics, - alignment-free methods, k-mers, hashing; locality sensitive hashing; min hashing; analysis; applications, - genome assembly and related problems, - motif search and discovery; statistical evaluation of motifs.
<p>Assessment/Exams:</p>	<p>You need a cumulative 50% of the points in the problem sets (in both theoretical and programming exercises) to be admitted to the oral exam.</p> <p>Dates for oral exams will be set by individual appointment, usually shortly after the end of the semester, or close to the beginning of the next semester (re-examination).</p>
<p>Literature:</p>	<p>Lecture slides and problem sheets are available on the website. Lecture recordings are available in the password protected area for registered students (CMS). In addition, the following books are recommended:</p> <p>Gonzalo Navarro, Mathieu Raffinot: <i>Flexible Pattern Matching in Strings</i>. Cambridge University Press.</p> <p>Enno Ohlebusch: <i>Bioinformatics Algorithms: Sequence Analysis, Genome Rearrangements, and Phylogenetic Reconstruction</i>. Oldenbusch Verlag, 2013.</p> <p>Dan Gusfield: <i>Algorithms on Strings, Trees and Sequences</i>. Cambridge University Press.</p>