

# Bioinformatics Algorithms

(Fundamental Algorithms, module 2)

**Zsuzsanna Lipták**

Masters in Medical Bioinformatics  
academic year 2018/19, II semester

Organisation

# Organisation

- Title of course: **Bioinformatics Algorithms**  
(Fundamental Algorithms, module 2)  
Master of Medical Bioinformatics (MB)  
**6 CFU** of a total of 12 CFU
- This course doubles as (*mutuato*)  
**Algorithms for Computational Biology**  
in the Master in Medical and Molecular Biotechnology (MMB)  
**6 CFU**

## Organisation (cont.)

- **course times:** Tue 12:30 - 14:30 (aula L), Thu 11:30 - 14:30 (aula A)
- **email:** `zsuzsanna.liptak@univr.it`  
Please include the course title and your name in the email
- **office:** CV 2, 1st floor, room 1.79
- **student hours:** Wed 10-12 (9:30-11:30?) and by appointment
- **webpage of course:**  
`http://profs.scienze.univr.it/~liptak/FundBA/`

## Organisation (cont.)

- **exam**: written and oral,  
admitted to oral only if you pass the written test
- **different exams** for students of MB and MMB
- There will be **two extra lectures** for students of MMB  
on computational complexity
- **takehome exercises** during term: will be discussed but not marked
- for Fundamental Algorithms **final grade** is 50% mod.1, 50% mod.2

## Organisation (cont.)

- **exam**: written and oral, admitted to oral only if you pass the written test
- **different exams** for students of MB and MMB
- There will be **two extra lectures** for students of MMB on computational complexity
- **takehome exercises** during term: will be discussed but not marked
- for Fundamental Algorithms **final grade** is 50% mod.1, 50% mod.2

**Questions?**

## Overview (tentative)

- Pairwise sequence analysis
  - Pairwise sequence alignment (global, local, other variants)
  - Pairwise alignment in practice: BLAST, Scoring matrices
  - String distances (edit distance, LCS distance,  $q$ -gram distance)

# Overview (tentative)

- Pairwise sequence analysis
  - Pairwise sequence alignment (global, local, other variants)
  - Pairwise alignment in practice: BLAST, Scoring matrices
  - String distances (edit distance, LCS distance,  $q$ -gram distance)
- Sequence assembly algorithms
  - Sanger shotgun sequencing: SCS (recap)
  - Sequencing with de Bruijn graphs

# Overview (tentative)

- Pairwise sequence analysis
  - Pairwise sequence alignment (global, local, other variants)
  - Pairwise alignment in practice: BLAST, Scoring matrices
  - String distances (edit distance, LCS distance,  $q$ -gram distance)
- Sequence assembly algorithms
  - Sanger shotgun sequencing: SCS (recap)
  - Sequencing with de Bruijn graphs
- Multiple sequence alignment
  - DP-algorithm, SP-score
  - Heuristic and approximation algorithms



# Overview (tentative)

- Pairwise sequence analysis
  - Pairwise sequence alignment (global, local, other variants)
  - Pairwise alignment in practice: BLAST, Scoring matrices
  - String distances (edit distance, LCS distance,  $q$ -gram distance)
- Sequence assembly algorithms
  - Sanger shotgun sequencing: SCS (recap)
  - Sequencing with de Bruijn graphs
- Multiple sequence alignment
  - DP-algorithm, SP-score
  - Heuristic and approximation algorithms
- Basics of Phylogenetics
  - distance-based data: UPGMA, Neighbor Joining
  - character-based data: Perfect Phylogeny, Small and Large Parsimony

# Overview (tentative)

- Pairwise sequence analysis
  - Pairwise sequence alignment (global, local, other variants)
  - Pairwise alignment in practice: BLAST, Scoring matrices
  - String distances (edit distance, LCS distance,  $q$ -gram distance)
- Sequence assembly algorithms
  - Sanger shotgun sequencing: SCS (recap)
  - Sequencing with de Bruijn graphs
- Multiple sequence alignment
  - DP-algorithm, SP-score
  - Heuristic and approximation algorithms
- Basics of Phylogenetics
  - distance-based data: UPGMA, Neighbor Joining
  - character-based data: Perfect Phylogeny, Small and Large Parsimony
- Introduction to string data structures
  - Basics of Suffix Trees and Suffix Arrays
  - Some applications

## Books

- **Enno Ohlebusch**: Bioinformatics Algorithms: Sequence Analysis, Genome Rearrangements, and Phylogenetic Reconstruction. Oldenbusch Verlag (2013) —recent, detailed, covers some but not all topics of this course, 3 copies in library
- **H.-J. Böckenhauer, D. Bongartz**: Algorithmic Aspects of Bioinformatics (2010)
- **V. Mäkinen, D. Belazzougui, F. Cunial, A.I. Tomescu**: Genome-Scale Algorithm Design. Cambridge University Press (2015)—very recent, advanced
- **Neil C. Jones and Pavel A. Pevzner**: An Introduction to Bioinformatics Algorithms (2004)—3 copies in library
- **David M. Mount**: Bioinformatics: Sequence and Genome Analysis (2004)—biologically oriented book, detailed, not always sufficiently algorithmic
- **João Setubal, João Meidanis**: Introduction to Computational Molecular Biology (1997)—my old favorite but a bit dated, 1 copy in library
- **Dan Gusfield**: Algorithms on Strings, Trees, and Sequences (1997)—the bible of string algorithms, a bit dated now
- **Joseph Felsenstein**: Inferring Phylogenies (2004)—important book on phylogenetics, very understandably written
- **Cormen, Leiserson, Rivest (& Stein)**: Introduction to Algorithms (different editions, 1990-onwards)—the bible of algorithms, a must-have for anyone interested in algorithms (buy second hand, old editions are also fine)