

## Summary

Course "Discrete Biological Models" (Modelli Biologici Discreti)

Zsuzsanna Lipták

Laurea Triennale in Bioinformatica  
a.a. 2014/15, fall term

Goals of this course:

- learn about some **discrete models** used in bioinformatics: models that use **strings (= sequences)**, **graphs**, **integer matrices**, **permutations**, ...
- **recognize** situations where similar models can be applied, and **develop** simple models for given phenomena
- improve knowledge of **discrete mathematics** (combinatorics, strings, graphs)

For this we will need to:

- study in detail some **discrete models used in biology**
- study **basic discrete mathematics** (combinatorics, counting, fundamental sequences, strings, graphs, trees)
- understand **NP-completeness**

2 / 6

### What did we do?

- What is **modeling**?
  - Some basic **discrete maths**: basic counting principles, binomial coefficients, graphs
  - **NP-completeness**
- (continued)

3 / 6

### What did we do?

- Biological problem 1 – **Sanger Sequencing** (Fragment Assembly, classical)  
**models/techniques**:
    - SCS (simplistic, **NP-hard**) - **overlap graph** (SCS = **heaviest Hamiltonian path**);  
Greedy algorithm (**efficient, approximate**)
    - Reconstruction (more complex, **NP-hard**)
    - Multicontig (even more complex, **NP-hard**)
- (continued)

4 / 6

### What did we do?

- Biological problem 2: **SBH (Sequencing by Hybridization)**  
**models/techniques**: **de Bruijn graph**, **Euler path/tour**, **efficiently solvable** (Hierholzer's algorithm)
  - Biological problem 3:  
**NGS-Sequencing: Fragment Assembly for Short Reads**  
**models/techniques**: **de Bruijn graph**, **Euler path/tour**, **efficiently solvable** (Hierholzer's algorithm);  
N.B.: Problem needed adjustment before applying this technique!
- (continued)

5 / 6

### What did we do?

- Biological problem 4: **Phylogenetic Reconstruction**  
**models/techniques**: **phylogenetic trees** (different types)
  - distance data: **path metric = input distances**; for certain inputs **efficiently solvable**
  - character data: **PP**, **Maximum Parsimony**, **NP-hard**

6 / 6