

1 General

The specialisation Bioinformatics includes applied biology (especially genetics and molecular biology) and the bioinformatics area of applied computer science. It is offered in cooperation with the Faculty of Biology and Psychology.

2 Lectures

Bioinformatics

- ▶ Bioinformatik in einer forschungsbezogenen Projektarbeit
- ▶ Systembiologie (T. Beißbarth)
- ▶ Seminar Bioinformatik
- ▶ Algorithmen der Bioinformatik II (P. Meinicke, B.Morgenstern)
- ▶ Grundlagen der Biostatistik mit R
- ▶ Data Mining in der Bioinformatik (P. Meinicke)

Biology (selection)

- ▶ Biochemie
- ▶ Mikrobiologie
- ▶ Allgemeine Entwicklungs- und Zellbiologie
- ▶ Genetik und mikrobielle Zellbiologie
- ▶ Neurobiologie
- ▶ Evolution und Systematik der Tiere

3 Research Example: Sequence Analysis

Traditional methods for comparative sequence analysis and phylogeny reconstruction are based on pairwise and multiple sequence alignment. These methods reach their limits if whole genomes are to be aligned.

Problems with sequence alignments:

- ▶ Computationally difficult (optimal MSA: NP hard)
- ▶ Homologous regions may be rearranged or duplicated
- ▶ Lateral gene transfer

Consequently, alignment-free methods are increasingly used for genome comparison.

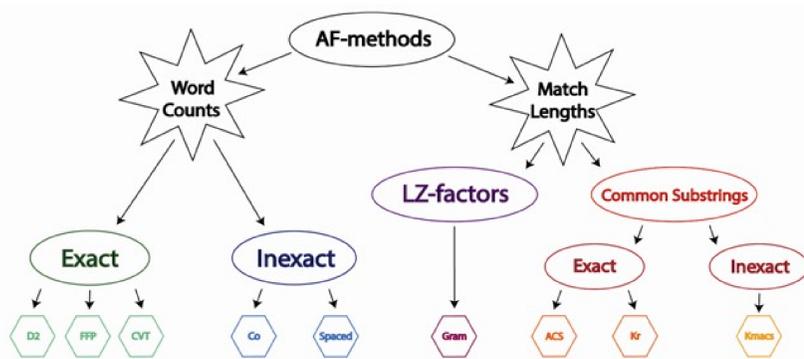


Figure 1: Classification of alignment-free methods

Main application:

- ▶ Phylogeny reconstruction

Procedure:

- ▶ Pairwise distance calculation (alignment-free)
- ▶ Calculate tree based on these distances (e.g. Neighbour-Joining)

Results:

- ▶ Less accurate than alignment-based phylogeny
- ▶ But much faster!

B. Morgenstern's group: development of alignment-free methods based on spaced-words [2]. Example:

```
Seq1      A C G A A C T T G A C A T T G A C C G G A
Seq2    G A C A C C G G A A C A T G C C G T T T G
Pattern   1 0 0 1 1 0 1 0 1 1
```

4 Research Example: Computational Metagenomics

Applying machine learning and data mining techniques to analyze sequence data from microbial communities. Basis: large sequence data sets from clinical samples (human microbiome) or environmental samples (marine, soil, ...).

Questions:

- ▶ What is the taxonomic composition of the community?
- ▶ What is the metabolic potential of the community?
- ▶ How does the community behave under changing external conditions?
- ▶ How do different organisms interact?

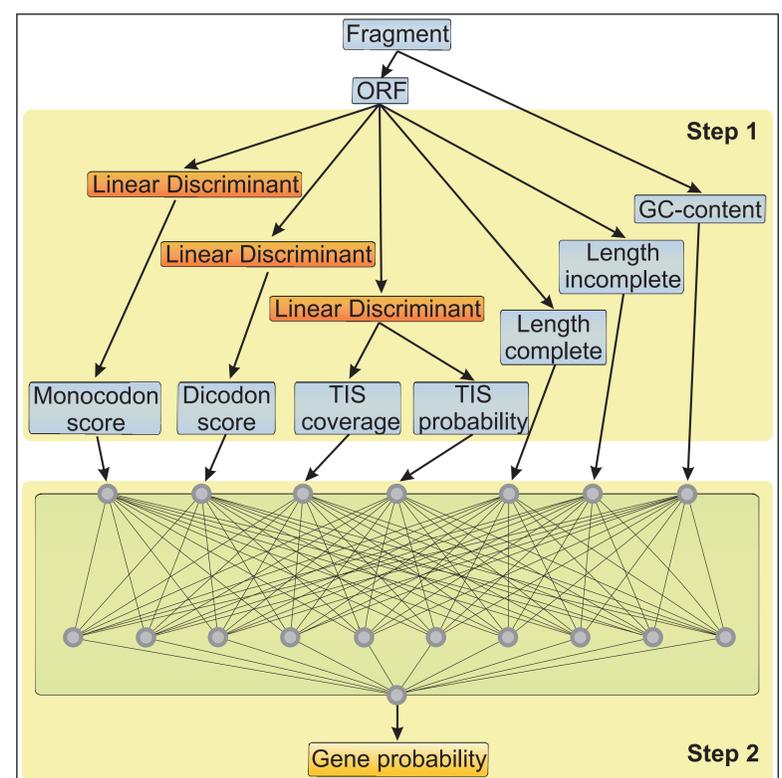


Figure 2: Neural network for metagenomic gene prediction

Problems:

- ▶ Mixed species data (unknown origin)
- ▶ Short genomic fragments (incomplete information)
- ▶ Really big data (computational challenge!)

5 References

- [1] Heiner Klingenberg and Peter Meinicke. How to normalize metatranscriptomic count data for differential expression analysis. *PeerJ*, 5:e3859, October 2017.
- [2] Chris-Andre Leimeister, Salma Sohrabi-Jahromi, and Burkhard Morgenstern. Fast and accurate phylogeny reconstruction using filtered spaced-word matches. *Bioinformatics*, 33(7):971, 2017.
- [3] Kathrin P. Ahauer, Bernd Wemheuer, Rolf Daniel, and Peter Meinicke. Tax4fun: predicting functional profiles from metagenomic 16s rRNA data. *Bioinformatics*, 31(17):2882–2884, 2015.