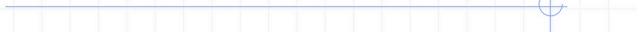


Introduction to Bioinformatics

Thomas Berlage
WS 2014/15



About me

(1) Computer scientist by training

- Worked on engineering interactive software for my PhD
- Then switched to focus on biomedical applications

(2) Institute director at Fraunhofer FIT

- Located in Sankt Augustin near Bonn
- Heading the the department of Life Science Informatics
- Working for customers in industry and research

(3) Working on high content analysis

- Automated optical instruments
- Imaging cells and tissue extensively
- Automated image and data analysis
- Data management
- Visual data mining

0-1

Lecture notes

(1) Bonn:

- Send mail with subject
 - ♦ Bioinformatics
- to
 - ♦ thomas.berlage@fit.fraunhofer.de
- to be added to a BSCW workspace

(2) Aachen

- L2P
- You need to register to the lecture (should still be possible)

(3) Podcast

- I am trying to deliver the whole lecture as a podcast to be used together with the slides so that there is more time for discussion during the lecture

0-2

Exam

(1) Written exam

- Some multiple choice
- Most need written answers (English or German accepted)

(2) Questions both addressing computer science and biology

- Getting one half right lets you pass
- Good students need to master both sides
- One side will be difficult for you (depending on your background)

(3) Date

- At the end of the lecture
- Will be announced
- In Aachen, you should register via VZPA to make my life easier

0-3

Example: Human Genome Project

(1) World-wide project (1990-2005)

- to establish complete sequence (3 billion letters) of the human genome
- to identify functional areas (around 35.000 genes)

(2) Tasks for bioinformatics

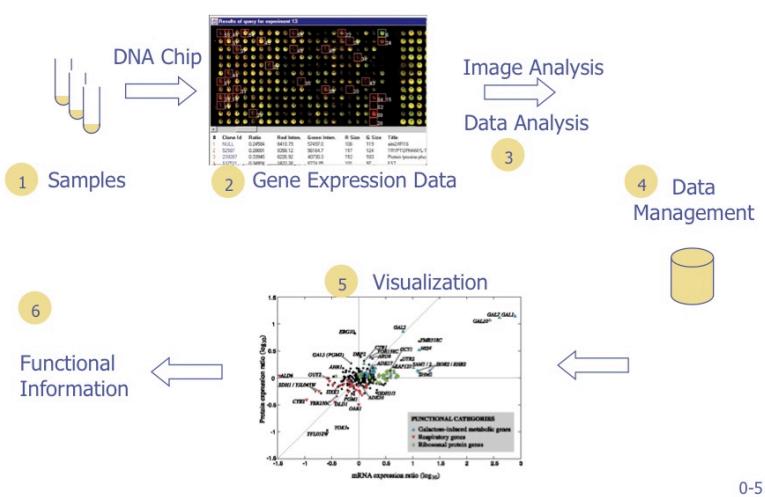
- Store and annotate the collected information
- Assemble partial sequences (overlapping areas)
- Find characteristic sequences indicating genes
- Compare variations between individuals
- Compare sequences with other organisms (evolutionary conserved genes)

(3) Findings

- Humans have only twice as many genes as worms!
- Not „one gene – one function“

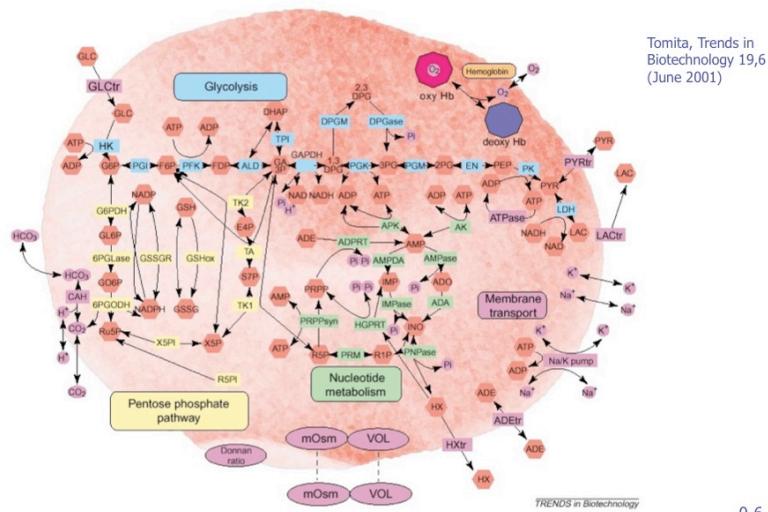
0-4

Example: Gene Expression Analysis



0-5

Example: Cell Simulation (E-Cell)



0-6

What is bioinformatics?

(1) Algorithmic view

- Algorithms on DNA sequences
- Anything to do with genome databases
- Computational molecular biology

(2) Application view

- Computer as an instrument to understand biological systems
- Data analysis, data management, and computational modeling
- Primarily large-scale, molecular, systematic

(3) Market view

- Special software used in biotechnology and pharmaceutical industry
- Laboratory information systems, instrument control
- Management of clinical studies, knowledge management

0-7

Analysis Levels of Biological Systems

(1) Ecosystem

- Food chains, evolution

(2) Organism

- Development, functional decomposition

(3) Organs/Physiology/Tissue

- Specialization, metabolism

(4) Cells/Inter-Cell Communication

- Development, signaling

(5) Genome/Molecular Function

- Transkription, translation, transport, regulation

(6) Physical Chemistry

- Protein folding, docking

0-8

Complexity

(1) Quantitative

- 23.000 genes
- 1.000.000 proteins
- billions of molecules
- many reactions every second

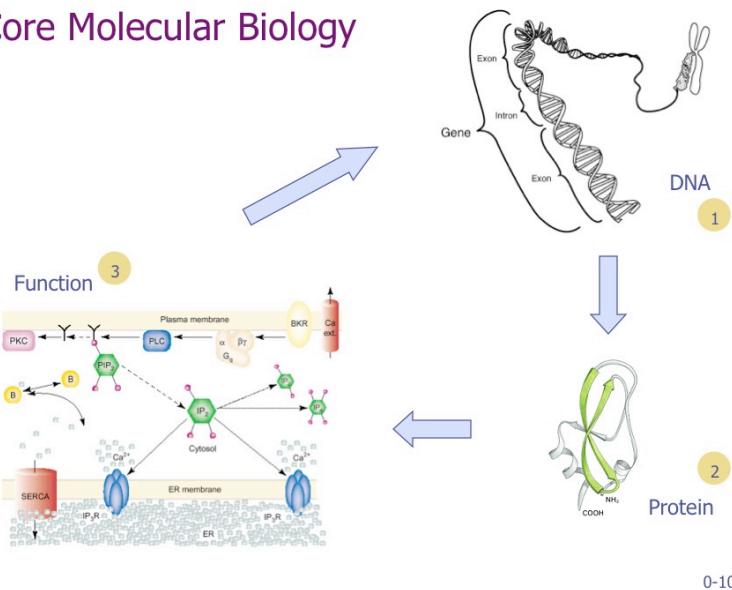
(2) Qualitative

- orders of magnitude between analysis levels
- hundreds of cell types
- development and specialization
- individual genome variation
- many species
- environmental effect and interaction

(3) Finally: What is consciousness?

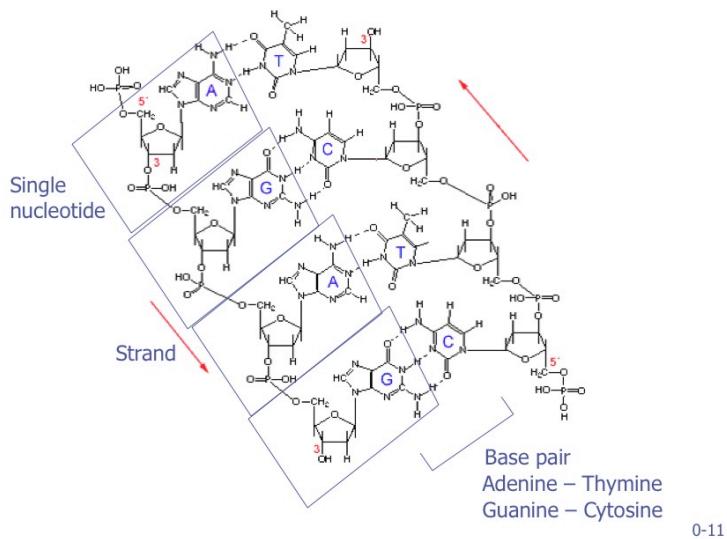
0-9

Core Molecular Biology



0-10

DNA



0-11

Proteins

- (1) Proteins are composed of 20 different amino acids
 - Single strand – thousands of amino acids
 - Sequential synthesis (from genetic code)
- (2) Proteins fold in characteristic ways
 - Flexible binding between adjacent amino acids
 - Positive or negative forces between nearby amino acids
 - This 3D structure is very important for protein function („docking“ of ligand molecules)
- (3) Protein sequence thus determines function

0-12

Overview

(1) Genome and Sequences

- DNA sequences
- Algorithms for sequence comparison
- Sequence databases
- Patterns and motifs
- Phylogenetic trees

(2) Proteins and Structures

- Protein structures
- 3D modelling
- Protein databases
- Protein structure analysis
- Protein structure prediction

(3) Protein Expression and Function

- DNA chip technology
- Gene expression analysis
- Clustering
- Proteomics

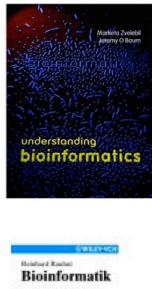
(4) Pathways and Systems

- Metabolic networks
- Pathway analysis
- Cell simulation

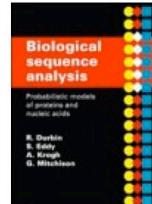
0-13

Books

(1) Zvelebil/Baum
Understanding Bioinformatics



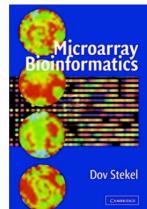
(2) Durbin/Eddy/Krogh/Mitchison
Biological sequence analysis



(3) Rauhut
Bioinformatik



(4) Stekel
Microarray Bioinformatics



1-14