MSc Bioinformatics
Course
Fundamentals of Bioinformatics

Lecture 1: Introduction

Centre for Integrative Bioinformatics VU (IBIVU)
Faculty of Exact Sciences / Faculty of Earth and Life Sciences
http://ibi.vu.nl, heringa@few.vu.nl, 87649 (Heringa), Room P1.28
MSc Bioinformatics & Systems Biology

- Major Bioinformatics
- Major Systems Biology
Overview of the BSB master

Year 1:
60 ec courses
• 42 ec compulsory
• 18 ec optional (up to 12 ec deficiency mending)

Year 2:
Internships (major/minor)
- Total 60 ec
Structure of the BSB master

Course work
- 4 courses (24 ec) compulsory for both BI & SB
- 3 further courses (18 ec) compulsory for BI
- 2 further courses (18 ec) compulsory for SB
- 3 optional courses (18 ects) for both BI & SB

Internships
- Total 60 ec (Major 33-42 ec, Minor 27-18 ec)
## Curriculum year 1

### Joint start Bioinformatics & Systems Biology

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>sept</td>
<td>oct</td>
<td>nov</td>
</tr>
</tbody>
</table>

**Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)**

- **Fundamentals of Bioinformatics** (6 ec) VU (+ UvA?)
- **Introduction to Systems Biology** (6 ec) VU (+ UvA?)
- **Algorithms in Sequence Analysis** (6 ec) VU
- **Systems Biology in Practice** (12 ec) VU + UvA
- **Biosystems Data Analysis** (6 ec) UvA

---

- **Joint subject BI & SB**
- **Decision moment**
- **Full schedule**
- **Bioinformatics**
- **Systems Biology**
- **Joint start Bioinformatics & Systems Biology**
Curriculum year 1

<table>
<thead>
<tr>
<th>Period 1</th>
<th></th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>sept</td>
<td>oct</td>
<td>nov</td>
<td>dec</td>
</tr>
<tr>
<td><strong>Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fundamentals of Bioinformatics (6 ec) VU (+ UvA?)</td>
<td></td>
<td>Algorithms in Sequence Analysis (6 ec) VU</td>
<td></td>
</tr>
<tr>
<td>Introduction to Systems Biology (6 ec) VU (+ UvA?)</td>
<td></td>
<td></td>
<td>Biosystems Data Analysis (6 ec) UvA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Systems Biology in Practice (12 ec) VU + UvA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Period 4</th>
<th></th>
<th>Period 5</th>
<th>Period 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>feb</td>
<td>mar</td>
<td>apr</td>
<td>may</td>
</tr>
<tr>
<td></td>
<td></td>
<td>jul</td>
<td>jun</td>
</tr>
<tr>
<td>Structural Bioinformatics (6 ec) VU</td>
<td></td>
<td>Bioinformatics of Large Systems (6 ec) VU (+ UvA?)</td>
<td>Academic Skills, Proposal &amp; Seminars (6 ec) VU + UvA</td>
</tr>
<tr>
<td>Basic Models of Biological Networks (6 ec) VU</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Curriculum Year 1

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
<th>Period 5</th>
<th>Period 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>sept</td>
<td>oct</td>
<td>nov</td>
<td>dec</td>
<td>jan</td>
<td>feb</td>
</tr>
<tr>
<td><strong>Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)</strong></td>
<td><strong>Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)</strong></td>
<td><strong>Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)</strong></td>
<td><strong>Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)</strong></td>
<td><strong>Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)</strong></td>
<td><strong>Bioinformatics and Systems Biology - Compulsory (24 + 18 ec)</strong></td>
</tr>
<tr>
<td>Fundamentals of Bioinformatics (6 ec) VU (+ UvA?)</td>
<td>Algorithms in Sequence Analysis (6 ec) VU</td>
<td>Biosystems Data Analysis (6 ec) UvA</td>
<td>Structural Bioinformatics (6 ec) VU</td>
<td>Bioinformatics of Large Systems (6 ec) VU (+ UvA?)</td>
<td>Proposal Writing and Seminar Series (6 ec) VU flexible*</td>
</tr>
<tr>
<td>Introduction to Systems Biology (6 ec) VU (+ UvA?)</td>
<td>Systems Biology in Practice (12 ec) VU + UvA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Optional Recommended Courses</strong></td>
<td><strong>Optional Recommended Courses</strong></td>
<td><strong>Optional Recommended Courses</strong></td>
<td><strong>Optional Recommended Courses</strong></td>
<td><strong>Optional Recommended Courses</strong></td>
<td><strong>Optional Recommended Courses</strong></td>
</tr>
<tr>
<td>Molecular Structures in Biology (6 ec) UvA</td>
<td>Advanced Modeling in Systems Biology (6 ec) VU</td>
<td>Computational Biology (6 ec) UvA</td>
<td>Programming in R (6 ec) VU</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Preparatory Bachelor Courses (assigned to address deficiencies; max 12 ec)</strong></td>
<td><strong>Preparatory Bachelor Courses (assigned to address deficiencies; max 12 ec)</strong></td>
<td><strong>Preparatory Bachelor Courses (assigned to address deficiencies; max 12 ec)</strong></td>
<td><strong>Preparatory Bachelor Courses (assigned to address deficiencies; max 12 ec)</strong></td>
<td><strong>Preparatory Bachelor Courses (assigned to address deficiencies; max 12 ec)</strong></td>
<td><strong>Preparatory Bachelor Courses (assigned to address deficiencies; max 12 ec)</strong></td>
</tr>
<tr>
<td>Calculus 1 (6ec) VU</td>
<td>Calculus 1 (3ec) VU</td>
<td>Inleiding Programmeren (6ec) VU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Biology of Cell I (3 ec) VU</td>
<td>Collective Intelligence (6ec) VU</td>
<td>Machine Learning (6 ec) VU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lineaire algebra voor BWI en N (6ec) VU</td>
<td>Lineaire algebra I (6ec) VU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probleemoplossers (3ec) VU</td>
<td>Probleemoplossers (3ec) VU</td>
<td>Moloculaire Celbiologie en Genetica (6ec) VU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biochemie I (3ec) VU</td>
<td>Biochemie II (3ec) VU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Optional Courses Other Masters (possibly in second year)</strong></td>
<td><strong>Optional Courses Other Masters (possibly in second year)</strong></td>
<td><strong>Optional Courses Other Masters (possibly in second year)</strong></td>
<td><strong>Optional Courses Other Masters (possibly in second year)</strong></td>
<td><strong>Optional Courses Other Masters (possibly in second year)</strong></td>
<td><strong>Optional Courses Other Masters (possibly in second year)</strong></td>
</tr>
<tr>
<td>Neural Networks (6 ec) VU</td>
<td>Stochastic Simulation (6 ec) UvA</td>
<td>Understanding Molecular Simulation (6 ec) UvA</td>
<td>Scientific Computing (6 ec) UvA</td>
<td>Data Mining Techniques (6 ec) VU</td>
<td>Complex System Simulation (6 ec) UvA</td>
</tr>
<tr>
<td>Evolutionary Computing (6 ec) VU</td>
<td>Computer Graphics (6 ec) VU</td>
<td>Molecular Microbial Physiology (6 ec) UvA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parallel Programming (6 ec) VU</td>
<td>Physical Biology of Cell II (3 ec) VU</td>
<td>Molecular Cell Physiology and Function (6 ec) VU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principles of Neuroscience (6 ec) VU</td>
<td>Genomes and Gene Expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Key:</strong></td>
<td><strong>First year:</strong></td>
<td><strong>Second year:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compulsory</td>
<td>42 ects are compulsory:</td>
<td>60 ects of projects:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bioinformatics Profile</td>
<td>- 24 ects are compulsory for all students</td>
<td>- major (max. 42 ects) must match profile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems Biology Profile</td>
<td>- 18 ects differentiate between the Bioinformatics and Systems Biology profiles</td>
<td>(Bioinformatics or Systems Biology)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommended Optional Courses</td>
<td></td>
<td>- minor (min. 18 ects)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplementary Courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optional Courses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*schedule of ‘Proposal Writing and Seminar Series’ is flexible and not limited to Period 6 in the first year
Further optional courses

**Computer Science**
- Artificial Intelligence
- Genetic algorithms
- Neural Networks

**Statistics**
- Statistical Genetics
- Statistical Models

**Biology**
- Epigenetics
- Molecular physiology

**Other interesting options**
- Principles of Neuroscience
- Physical biology of the cell
Overview Year 2

• two internships
  • Minor: 18-27 ects
  • Major: 33-42 ects

• internship hosting:
  • Universities, Research institutes, Companies, Academic hospitals
Internships sites to date

- Netherlands, a.o.:
  - VU (FEW/ FALW)
  - VUmc
  - CMBI
  - UMCU
  - UvA
  - UU
  - TNO
  - MRC-Holland

- LUMC
- Philips
- Organon (Merck)
- Agendia
- Keygene
- Baseclear
International internship sites to date

– Stockholm Bioinformatics Centre
– European Bioinformatics Institute, Hinxton, UK
– Deutsches Krebs Forchungs Zentrum, Heidelberg, Germany
– European Molecular Biology Laboratory, Heidelberg, Germany
– INRIA Strasbourg
– Sloan Kettering Foundation NYC, USA
– LONI laboratory UCLA, USA
– Oakland University, NZ
– RIKEN Tokyo, Japan
– UMC Mineapolis (Ann Arbour), USA
Fundamentals of Bioinformatics

- Course coordinator: Dr Anton Feenstra
### FoB Schedule

#### Lectures 9.00-11.00:

<table>
<thead>
<tr>
<th>Week</th>
<th>Day</th>
<th>Room</th>
<th>Topic, contents</th>
<th>Lecturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mon</td>
<td>P6.24</td>
<td>General: Modeling, optimality Evolution: mutations, selection, sex orthologs</td>
<td>JH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue</td>
<td>P6.24</td>
<td>Genomics (Genes, genomics data, codons, translation/transcription, microarray &amp; proteomics data, splicing, GO-PSI-blast?)</td>
<td>JH</td>
</tr>
<tr>
<td>2</td>
<td>Tue</td>
<td>P6.24</td>
<td>Genomics (Sequence alignment / Bioinformatics for Systems Biology, Smith/Waterman (DP), sequence DBs)</td>
<td>JH</td>
</tr>
<tr>
<td></td>
<td>Mon</td>
<td>P6.24</td>
<td>Ontologies &amp; GO</td>
<td>AG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Mon</td>
<td>P6.24</td>
<td>Protein Bioinformatics: aa’s, ss, folds, domains, disorder, TM, PDB</td>
<td>SA</td>
</tr>
<tr>
<td></td>
<td>Tue</td>
<td>P6.24</td>
<td>Protein Bioinformatics: protein interactions &amp; thermodynamics</td>
<td>AF</td>
</tr>
<tr>
<td>4</td>
<td>Mon</td>
<td>P6.24</td>
<td>Petri-net models</td>
<td>AF</td>
</tr>
<tr>
<td></td>
<td>Tue</td>
<td>P6.24</td>
<td>Secondary Structure prediction (including Machine Learning)</td>
<td>JH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Mon</td>
<td>P6.24</td>
<td>Repeat detection and/or Domain prediction</td>
<td>JH</td>
</tr>
<tr>
<td></td>
<td>Tue</td>
<td>P6.24</td>
<td>t.b.d. (Next Gen Sequencing? / Wrapping up?)</td>
<td>WP?/JH?</td>
</tr>
<tr>
<td>6-7</td>
<td></td>
<td></td>
<td>– project work –</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Fri</td>
<td>Plantage</td>
<td>presentations + questions</td>
<td>JH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Acronym Lecturer**

- JH Jaap Heringa
- AF Anton Feenstra
- SA Sanne Abeln
- ME Mohamed El-Kebir
- AG Andrew Gibson
- NB Nicola Bonzanni
- PB Punto Bawono
- WP Walter Pirovano
Deficiencies worked on during September-October

• Three defined areas: programming, mathematics, biology
• two out of three are addressed (working on all three is possible)
  – Mon/Tue: deficiency in programming (Perl)
    • 11-12am, 1-2pm (Rm P3.37) – Fundamentals of Bioinformatics
  – Wed/Thu: deficiency in Mathematics
    • 11-12am, 1-2pm – Introduction to Systems Biology
  – Fri: deficiency in Biology (UvA)
Project/practical work

**Project: 14.00-16.00**

- Practicals weeks 1-5
  - Intro
  - Find Protein(s), Pfam & GO terms
  - Find matching sequences (Blast) plus GO & Pfam entries
  - Find SCOP families
  - Scoring & Benchmarking
  - PSI-blast
- Project weeks 6-7
  - building on scripts & results from practicals
  - report, presentation
Some teachers in FoB

- Nicola Bonzanni – PhD (1/11/07)
- Anton Feenstra - UD (1/09/05) Course coordinator
- Mohammed El-Kebir – PhD (1/06/10)
- Sanne Abeln – UD (1/11/09)
- Jaap Heringa – Hgl (1/10/02)
- Punto Bawono – PhD (1/06/11)
- Andrew Gibson – PD AMC (Guest lecturer)
Examination and course marks

1. Oral or written exam  (40% weight)
   • General knowledge and overview
   • Possibly (if oral exam):
     o Questions about scientific papers
     o Discussion on scientific issue (use case)

2. Classes - deficiency  (30% weight)
   • One out of two deficiencies counts for course FoB, the other will count for the course ISB

3. Group (project) work  (30% weight)
   • Progress, report, project presentation
Gathering knowledge

- Anatomy, architecture

- Dynamics, mechanics

- Informatics
  - (Cybernetics – Wiener, 1948)
    (Cybernetics has been defined as the science of control in machines and animals, and hence it applies to technological, animal and environmental systems)
  - Genomics, bioinformatics, systems biology
Bioinformatics

Chemistry

Mathematics
Statistics

Computer Science
Informatics

Physics

Biology
Molecular biology

Medicine
Bioinformatics

“Studying informational processes in biological systems” (Hogeweg, early 1970s)

• No computers necessary
• Back of envelope OK

Applying algorithms with mathematical formalisms in biology (genomics)

Hazardous: biology and biological knowledge is crucial for making meaningful analysis methods!
How does information come in?

- Modeling a lion leaving its territory due to food shortage
- Prey density? What does this mean for the lion?
- \#prey/area?
- What else?
Why a model

“The proof of the pudding is in the eating”

A model should be an implementation of the current knowledge about a system, such that it can test that knowledge and lead to predictions that may be verified experimentally.
How to use a model

• it should be relevant
• “It should be as simple as possible (but not simpler” [Einstein])
• it should be realistic, although paradigm systems can be useful too
  – McCulloch-Pitts Neurons (1943)
  – McCulloch: “Don’t bite my finger but look where it’s pointing to”
• it should answer a question
  – asking a good question in science is perhaps more important than answering it
How to use a model

George E. P. Box

“.. all models are wrong, but some are useful”
How to use a model

1. Initially, it is very important to have a clearly defined aim. When the model assembled is able to satisfy the needs one should stop further modeling.

2. If the model is able to reproduce an experimental result and is additionally able to make predictions about yet unknown behaviours, the model suffices.

3. In general, the model should be able to reliably verify or falsify hypotheses made.
Why a model

1. To decide about model behaviour, one should be able to assess the model behaviour using some (experimental) standard of truth

2. This means that the state space of the model should be mapped onto the experimentally observed states
Haematopoietic (blood) stem cells

Differentiation tree
(50 cell types)

Observed expression states
\{10111001001\}
\{10010011010\}
\{00110000111\}
\{10010001101\}
\{00111001001\}

How to map observed state space onto model state space?

The state space and connections generated by the model
Models in Systems Biology

- are (still) predominantly based upon ordinary differential equations (ODEs)
- are often used to find equilibria in the system
- represent continuous models, unlike the blood stem cell model on the preceding slide that is discrete
- Remember that molecular biology is mostly discrete: it is about molecules interacting (although can deal with large numbers)
The starting point: the cell

Prokaryotes: Bacteria and Archaea bacteria
• DNA ‘floats’ freely in the cell
• No organelles (mitochondria, chloroplasts)

Eukaryotes: remaining species (plant, animal, fungus, protists, etc.)
• DNA in a compartment: nucleus, nucleolus
• Contain organelles (e.g. mitochondria, chloroplasts in plants, Golgi system, lysosomes)
A bacterial cell
A eukaryotic cell
Phylogenetics – the tree of life

Organisms are related through their evolutionary history, which can be traced using their genetic material.
The modern (interactive) tree of life

One of the main issues was cleaning up the tree by filtering out horizontally transferred genes in order to base the phylogeny on orthologous genes only.

**Letunic I** and Bork P (2007) Bioinformatics 23(1):127-8 *Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation*
Multicellular organisms:
Development of a zygote into a mature organism: many questions remain!
What makes a biological species: how are differences generated and what are the consequences of these differences?

- What is causing the difference between species? How do species arise?

- What is causing the difference between members of a population?
What makes us human?

Look at our nearest neighbor: Chimpanzee

- Differences in gene expression
- Positive selection for certain genes (e.g. HAR1 in the brain)
- Gene duplications
- Differences in ‘non-coding’ sequences
- Several other events

The human brain and the HAR1 gene

And the award for the fastest-evolving piece of human DNA goes to...

Out of 35,000 non-coding genes, 49 candidate segments have evolved a lot in the human lineage. The most drastically altered of all is a segment dubbed HAR1 (for human accelerated region 1). It is 118 base pairs long. Chimpanzees and chickens, separated by over 300 million years, carry versions of HAR1 that are identical except for two base pairs. In humans, on the other hand, 18 out of 118 base pairs have changed since we split from chimps.

Human cells make RNA molecules out of the HAR1 segment. Specifically, brain cells express HAR1 in the cortex, the hippocampus, and certain other regions. The sequence of HAR1 suggests that an RNA molecule produced from it would be stable enough to carry out some important job, such as regulating the activity of protein-coding genes. HAR1 probably plays several roles. It is not just active in the adult brain, but in development-guiding cells in the fetus.

HAR1, is part of a novel RNA gene (HAR1F) that is expressed specifically in Cajal–Retzius neurons in the developing human neocortex from 7 to 19 gestational weeks, a crucial period for cortical neuron specification and migration. HAR1F is co-expressed with reelin, a product of Cajal–Retzius neurons that is of fundamental importance in specifying the six-layer structure of the human cortex. HAR1 and the other human accelerated regions provide new candidates in the search for uniquely human biology.

However, HAR1 is also active in the ovary and testis of adult humans. Perhaps selection has acted on HAR1 in connection with reproduction, rather than with thought, although expression of HAR1 is far smaller in the sex cells than in the brain. Still, it’s a strange point that may be worth raising at your next party: we have genes that are only active in our brains and sex cells.

The human brain and the HAR1 gene

Recent evidence points at biased gene conversion (BGC) to explain the divergence of human accelerated region 1.

- BGC speeds up the rate of evolution in certain genes. This process increases the rate at which certain mutations spread through a population, regardless of whether they are beneficial or harmful.
- BGC is thought to be strongest in regions of high recombination, and can cause harmful mutations to spread through populations. Many of the genetic changes leading to human-specific characters may be the result of the fixation of harmful mutations. This contrasts the traditional Darwinistic view that they are the result of natural selection in favour of adaptive mutations.

Gene conversion is a process by which DNA sequence information is transferred from one DNA helix (which remains unchanged) to another DNA helix, whose sequence is altered. It is one of the ways a gene may be mutated. Gene conversion may lead to non-Mendelian inheritance.
Gene Conversion

• Conversion of one allele to the other due to base mismatch repair during recombination:
  – if one of the four strands during meiosis pairs up with one of the four strands of a different chromosome, as can occur if there is sequence homology, mismatch repair can alter the sequence of one of the chromosomes, so that it is identical to the other.
  – Gene conversion can result from the repair of damaged DNA as described by the Double Strand Break Repair Model. Here a break in both strands of DNA is repaired from an intact homologous region of DNA. Resection (degradation) of the DNA strands near the break site leads to stretches of single stranded DNA that can invade the homologous DNA strand. The intact DNA can then function as a template to copy the lost DNA.

• Gene conversion acts to *homogenize* the DNA sequences composing the gene pool of a species. Over time, gene conversion events lead to a homogenous set of DNA sequences, both for allelic forms of a gene and for multi gene families.
  – Interspersed repeats act to *block* gene conversion events, thus catalyzing evolution of new genes and species.

• Gene conversion has medical implications:
  – Gene conversion resulting in mutation of the CYP21A2 gene is a common underlying genetic cause of congenital adrenal hyperplasia.
  – Somatic gene conversion is one of the mechanisms that can result in familial retinoblastoma, a congenital cancer of the retina.
  – It is theorized that gene conversion may play a role in the development of Huntington's Disease.

• Conversely, biased gene conversion leads to accelerated mutation rates (see preceding slide). It happens primarily in recombination-prone areas.
Important questions in biology are dealing with the decoding of the ‘information’ that resides in the genetic material.

How can this……

DNA: Genotype

…lead to this?

Phenotype
What is Genomics

**Genomics** is a discipline in genetics concerning the study of the genomes of organisms. The field includes intensive efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping efforts. The field also includes studies of intragenomic phenomena and other interactions between loci and alleles within the genome.

- In contrast, the investigation of the roles and functions of single genes is a primary focus of molecular biology or genetics and is a common topic of modern medical and biological research. Research of single genes does not fall into the definition of genomics unless the aim of this genetic, pathway, and functional information analysis is to elucidate its effect on, place in, and response to the entire genome's networks.

A commonly used wider definition of genomics is that it involves the study of all the genes of a cell, or tissue, at the DNA (genotype), mRNA (transcriptome), or protein (proteome) levels.

Adapted from Wikipedia
Bioinformatics is closely associated with genomics

• The aim is to solve the genomics information problem
• Ultimately, this should lead to biological understanding how all the parts fit (DNA, RNA, proteins, metabolites) and how they interact (gene regulation, gene expression, protein interaction, metabolic pathways, protein signalling, etc.)
• Genomics will result in the “parts list” of the genome
• Bioinformatics aims at understanding the role of each of the components
Wrapping up

- General information about the master
- General information about the FoB course
- The role of information
- The importance and role of a model
- The cell: global properties
  - Prokaryotes (bacteria and archaea) and eukaryotes
  - Phylogeny
- Speciation and cell differentiation
- Genomics
  - Gene conversion / biased gene conversion
  - Accellerated evolution (HAR1)