Биоинформатика: предмет и методы. Биологическая информация. Ресурсы.

Что такое биоинформатика?
“Bioinformatics is the science of managing, mining, and interpreting information from biological data” *

*- http://bio.informatics.iupui.edu/biokdd10/
**BIOINFORMATICS IN THE 21st CENTURY**

Bioinformatics will be at the core of biology in the 21st century. In fields ranging from structural biology to genomics to biomedical imaging, ready access to data and analytical tools are fundamentally changing the way investigators in the life sciences conduct research and approach problems.

Complex, computationally intensive biological problems are now being addressed and promise to significantly advance our understanding of biology and medicine. No biological discipline will be unaffected by these technological breakthroughs.

workshop in White House, 1999
Pre-genomic → Post-genomic era

Science

The Human Genome

Nature

Nuclear fission
Five-dimensional energy landscapes

Seafloor spreading
The view from under the Arctic ice

Career prospects
Sequence creates new opportunities
# Bioinformatics analysis for ‘omics’ era in biology

<table>
<thead>
<tr>
<th>‘omics’</th>
<th>Object for study</th>
<th>Types of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics</td>
<td>Genome (full set of genes)</td>
<td>Genome assembly, gene finding, gene structure analysis, regulators identification, finding of sites of regulators binding with DNA</td>
</tr>
<tr>
<td>Transcriptomics</td>
<td>Transcriptome (full set of transcripts: mRNAs, rRNAs, tRNAs, ncRNAs)</td>
<td>Analysis of gene expression, RNA processing, splicing isoforms, RNA structure, ncRNAs functions</td>
</tr>
<tr>
<td>Proteomics</td>
<td>Proteome (entirety of proteins)</td>
<td>Proteome analysis, identification of proteins, analysis and prediction of protein structure and interactions</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>Metabolome (set of metabolites, regulators)</td>
<td>Small molecules identification, analysis of their transformations</td>
</tr>
</tbody>
</table>
All are in agreement about certain pivotal events that were true milestones:

- the double-helix model of DNA,
- The determination of the amino acid sequence of a protein, and
- the conceptual linking of DNA sequences and protein sequences

Sequences, the simple order of individual units in biological polymers, are at the heart of bioinformatics, and the search for relationships among them and the reconstruction of their histories has arguably proved the most informative of biological inquiries

Russell F. Doolittle
Department of Chemistry & Biochemistry, University of California
Information flows from DNA to RNA to proteins.

DNA (information storage)

3’ A A T G T G C C C G 5’

TRANSCRIPTION

mRNA (information carrier)

5’ U U A C A C G G G C 3’

Codon Codon Codon

TRANSLATION

Proteins (active cell machinery)

Leucine Histidine Glycine

http://www.uic.edu/classes/bios/bios100/lectures/proteins.htm
Развитие биохимических методов как основа исследования нуклеиновых кислот и белков

Разработка методов секвенирования (определение последовательности) НК и белков

→ накопление большого числа (тыс, млн) секвенированных последовательностей

→ потребность компьютерных подходов для их хранения (БД) и анализа (методы выравнивания, поиск эволюционных связей и функциональных мотивов, поиск гомологов в БД большого объема)
Frederick Sanger (born 1918) is a British biochemist who was twice the recipient of the Nobel Prize for Chemistry, the only person to have been so.
The Nobel Prize in Chemistry 1958 was awarded to **Frederick Sanger** "for his work on the structure of proteins, especially that of insulin".

The Nobel Prize in Chemistry 1980 was divided, one half awarded to Paul Berg "for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA", the other half jointly to Walter Gilbert and **Frederick Sanger** "for their contributions concerning the determination of base sequences in nucleic acids".
http://www.sanger.ac.uk/
Dr. Margaret Belle (Oakley) Dayhoff (1925 –1983) was an American physical chemist and a pioneer in the field of bioinformatics.

Dayhoff was a professor at Georgetown University Medical Center and a noted research biochemist at the National Biomedical Research Foundation where she pioneered the application of mathematics and computational methods to the field of biochemistry.

Dayhoff went on to pioneer the development of programmable computer methods for use in comparing protein sequences and deriving their evolutionary histories (in other words, discerning homologies) from their sequence alignments.
In order to reduce the size of the data files which describes the sequence of amino acids in a protein, Dr. Dayhoff shortened the three letter representations to single letter code.
She initiated the collection of protein sequences in the **Atlas of Protein Sequence and Structure**, a book collecting all known protein sequences that she published in 1965. It was subsequently republished in several editions.

This led to the **Protein Information Resource database** of protein sequences, which was developed by her group. It and the parallel effort by Walter Goad which led to the GenBank database of nucleic acid sequences are the **twin origins of the modern databases of molecular sequences**.
Celebrating 30 years of Bioinformatics

A symposium in the honor of Prof. dr. Paulien Hogeweg

Thirty years ago Paulien Hogeweg and Ben Hesper coined the term Bioinformatics [1-3] “for the study of informatic processes in biotic systems”. To celebrate this and Paulien’s 65th birthday we organize a symposium on Friday 19 December 2008 in the Boothzaal of the Utrecht University library.
As cost of sequencing halves every 12 months.
- cf Moore's Law

The Human genome project:
- 13 years.
- 23 labs.
- $500 Million.

A Human genome today:
- 3 days.
- 1 machine.
- $10,000.
- Large centres are now doing studies with 1000s and 10,000s of genomes.

Changes in sequencing technology are going to continue this trend.
- “Next-next” generation sequencers are on their way.
- $500 genome is probable within 5 years.
August 2007: 3.5 Gbases/month.
Jan 2010: 4 Tbases/month.
1000x increase in sequencing output.
In August 2007, total size of genbank was 200 Gbases.

Managing Growth

We have exponential growth in storage and compute

- Storage/compute doubles every 12 months.
  - 2009 ~7 PB raw

Gigabase of sequence ≠ Gigabyte of storage.

- 16 bytes per base for sequence data.
- Intermediate analysis typically need 10x disk space of the raw data.

Moore's law will not save us.

- Transistor/disk density: $T_d=18$ months
- Sequencing cost: $T_d=12$ months
International Nucleotide Sequence Database Collaboration

- GenBank (USA)
- EMBL-Bank (Europe)
- DDBJ (DNA Databank of Japan)

http://www.insdc.org/

International Nucleotide Sequence Database Collaboration

- The International Nucleotide Sequence Databases (INSDC) have been developed and maintained collaboratively between DDBJ, ENA, and GenBank for over 18 years.
- The INSDC advisory board, the International Advisory Committee, is made up of members of each of the databases' advisory bodies. At their most recent meeting, members of this committee unanimously endorsed and reaffirmed the existing data-sharing policy of the three databases that make up the INSDC, which is stated below.
- Individuals submitting data to the international sequence databases should be aware of INSDC policy.

How to submit data

- For full details of how to submit data to the databases, please select a collaborating database:
  - DDBJ, ENA, GenBank
- The INSDC Feature Table Definition Document is available here.
The scientific disciplines and expertises

- Bioinformatics
- Informatics
- biostatistics
- Biomath
- Medical Informatics
- Epidemiology
- Image analysis

Knowledge
Предпосылки появления биоинформатики

1) развитие биохимических и биофизических методов исследования (разделение и очистка молекул, секвенирование, микрочипы, кристалллография и т.д.)

2) Развитие компьютерной техники (мощность и объем) и методов программирования

3) Развитие сетевых технологий (интернет, ГРИД-технологии)
### Carriers of linear biological information

<table>
<thead>
<tr>
<th>Object</th>
<th>Monomers</th>
<th>Types of monomers</th>
<th>Approx. length (nbases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>(deoxyribo)-nucleotides</td>
<td>4 (AGCT)</td>
<td>10,000 -100,000</td>
</tr>
<tr>
<td>Transcript</td>
<td>(ribo)-nucleotides</td>
<td>4 (AGCU)</td>
<td>?&lt;50 -10,000</td>
</tr>
<tr>
<td>Protein</td>
<td>Amino acids</td>
<td>20 + modified</td>
<td>50-1,000 or &gt;</td>
</tr>
</tbody>
</table>
1D biological Information:
sequences

Gene of hemopexin (about 11,000 nb)

mRNA (1652 nb)

Protein (462 aa)
Типы биологической информации:

3D = пространственная структура (в осн. белков, но также РНК) – закодирована в линейной, основа для реализации функций;
Levels of protein structure: primary (sequence), secondary, tertiary, quaternary
Типы биологической информации:

4D = реакции, протекающие во времени,
сети взаимодействия молекул
4D biological Information: networks, metabolites flows
Разработка основных направлений биоинформатики

- Создание баз данных биологической информации
- Развитие алгоритмов и создание программ (инструментов)
- Анализ и интерпретация биологических данных

Интернет-пользователи (биологи, фармацевты, медики и т.д.)

Мировые центры секвенирования

Программисты
Применение биоинформатики для анализа биомолекул

- Gene finding
- Phylogenetic analysis
- Function prediction
- Structure modeling
- Ligand docking and drug design
- Physiological effect modeling
- Etc.
Области биологии, в которых проводится биоинформационный анализ:

- Области биологии, в которых проводится биоинформационный анализ:

**FEMS Microbiology Letters**

Volume 280 Issue 2, Pages 160 - 168
Published Online: 31 Jan 2008
© 2008 Federation of European Microbiological Societies. Published by Blackwell Publishing Ltd. All rights reserved

**Abstract | Full Text: HTML, PDF (Size: 196K) | Supporting Information | E-mail**

**RESEARCH LETTER**

Expression of the phytochrome operon in *Pseudomonas*

- Expression of the phytochrome operon in *Pseudomonas*
sigma factor RpoS

Katalin Barkovits1, Andrea Harms2, Cornina Benkartek2, James L. Smart1,3 & Ingo Hausmans3

1 Physiology of Microorganisms, Ruhr-University Bochum, Bochum, Germany; 2 Institute of Biology, University of Twente, Enschede, The Netherlands; and 3 Department of Biological Sciences, University of Texas, Arlington, Texas

**Correspondence:** Nicole Frankenberg-Dinkel, Physiology of Microorganisms, Ruhr-University Bochum, 44780 Bochum, Germany; Tel.: +49 234 32 23101; fax: +49 234 32 14620; e-mail: nicole.frankenberg@rub.de

**Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses**

Peter Simmonds1,*, Ioannis Karakasiliotis2, Dalan Bailey2,
Какие преимущества дает для современного биолога знание биоинформатики?
Why should you study bioinformatics?

Bioinformatics is an interesting, and rapidly advancing field with many diverse challenges. For instance, you can become an expert in

- Finding new genes and drug targets
- Personalized treatment of cancer or other disease
- Analysis of ancient DNA
- Heuristic sampling methods
- Predicting protein structure or conformation

There is a rapidly increasing need for bioinformatics expertise in both academia and industry.

There are great possibilities for continuing as a PhD student after the master program
Why should you study bioinformatics?

Bioinformatics is an interesting, and rapidly advancing field with many diverse challenges. For instance, you can become an expert in:

- Finding new genes and drug targets
- Personalized treatment of cancer or other disease
- Analysis of ancient DNA
- Heuristic sampling methods
- Predicting protein structure or conformation

There is a rapidly increasing need for bioinformatics expertise in both academia and industry.

There are great possibilities for continuing as a PhD student after the master program.
Журналы в области биоинформатики, компьютерной биологии и omics-исследований
## Impact factor in bioinformatics field

<table>
<thead>
<tr>
<th>Journal</th>
<th>Publisher</th>
<th>IF 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Briefings in Bioinformatics</td>
<td>Oxford journals</td>
<td>9.283</td>
</tr>
<tr>
<td>Nucleic Acids Research</td>
<td>Oxford journals</td>
<td>7.836</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>Oxford journals</td>
<td>4.877</td>
</tr>
<tr>
<td>DNA Research</td>
<td>Oxford journals</td>
<td>4.754</td>
</tr>
<tr>
<td>The European Journal of Human Genetics</td>
<td>Nature Publishing Group</td>
<td>4.380</td>
</tr>
<tr>
<td>DNA and Cell Biology</td>
<td>Mary Ann Liebert</td>
<td>2.159</td>
</tr>
<tr>
<td>Journal of Computational Biology</td>
<td>Mary Ann Liebert</td>
<td>1.694</td>
</tr>
<tr>
<td>Computers in Biology and Medicine</td>
<td>ELSEVIER</td>
<td>1.234</td>
</tr>
</tbody>
</table>
• **The International Society for Computational Biology** (ISCB) serves over 2500 members from nearly 70 countries around the world by addressing scientific policies, providing access to high quality publications, organizing meetings, and serving as a portal to information about training, education, employment and news from related fields.
Join us in Basel for ECCB 2012!

ECCB'12 - the European Conference on Computational Biology - is the key European computational biology event in 2012 uniting scientists working in a broad range of disciplines, including bioinformatics, computational biology, biology, medicine, and systems biology. Participation at ECCB'12 will be the prime opportunity to learn about cutting-edge research in computational biology and bioinformatics and to network with other members of our community.

ECCB'12 will be organized on 9-12 September 2012 in association with the 10th [BC]² Basel Computational Biology conference - the annual scientific symposium of the SIB Swiss Institute of Bioinformatics.

Important Dates:

Paper Submission Deadline: Nov. 28 2011
Notification of Acceptance: Dec. 29, 2011
Pre-registration and Camera-ready Manuscripts: January 20, 2012
Why should you study bioinformatics?

Bioinformatics is an interesting, and rapidly advancing field with many diverse challenges. For instance, you can become an expert in:

- Finding new genes and drug targets
- Personalized treatment of cancer or other disease
- Analysis of ancient DNA
- Heuristic sampling methods
- Predicting protein structure or conformation

There is a rapidly increasing need for bioinformatics expertise in both academia and industry.

There are great possibilities for continuing as a PhD student after the master program.
Bioinformatics jobs

Data

Local storage

Processing/curation

Presentation

Extract value

Database administration
Hardware support
IT knowledge essential
Might have scientific knowledge

Knowledge of Perl, Oracle, Unix
Understanding of biology/drug discovery

Knowledge of JAVA, HTML
Good scientific knowledge likely

Data miners
Understanding of biology/drug discovery essential
http://www.hopesandfears.com/hopesandfears/all/newprof/115933-novaya-professiya-bioinformatik

ПРОФЕССИЯ БУДУЩЕГО

Биоинформатик
R&D = research & development

In general, R&D activities are conducted by specialized units or centers belonging to a company, or can be out-sourced to a contract research organization, universities, or state agencies.

In the context of commerce, "research and development" normally refers to future-oriented, longer-term activities in science or technology, using similar techniques to scientific research but directed toward desired outcomes and with broad forecasts of commercial yield.
Research Groups & Projects

- **Biometaphorical Computing**: Neuroimaging, Neuroanalysis, and Modeling the Brain.

- **Blue Gene**: Science with Petascale to Exascale Computing.

- **Computational Genomics**: Making sense of the Data Explosion in Genomics.

- **DNA Transistor**: Nanotechnology for DNA Sequencing.

- **Functional Genomics & Systems Biology**: Transcriptomics, Reverse Engineering of Networks, Multiscale modeling of Biological Systems.

- **Genographic**: Human Migratory History from Genetic Data. Collaboration with National Geographic Society.
Why should you study bioinformatics?

Bioinformatics is an interesting, and rapidly advancing field with many diverse challenges. For instance, you can become an expert in:

- Finding new genes and drug targets
- Personalized treatment of cancer or other diseases
- Analysis of ancient DNA
- Heuristic sampling methods
- Predicting protein structure or conformation

There is a rapidly increasing need for bioinformatics expertise in both academia and industry.

There are great possibilities for continuing as a PhD student after the master program.
http://www.binf.ku.dk/programmes/Master
Master's in Bioinformatics at Stockholm Bioinformatics Centre, Stockholm University

The master's program in bioinformatics is aimed at students with a bachelor degree in natural science with a deep interest to understand biology using computational methods. With the rapid growth of high throughput methods in DNA sequencing and functional genomics, the amount and complexity of biological data poses very daunting challenges. Computational approaches are however under rapid development to meet these challenges.

Our master's program in bioinformatics is focused on techniques for understanding biology. Therefore several of the obligatory courses are aimed at understanding the molecular details of life. All courses are given by professors that are performing world class research, in comparative genomics, protein bioinformatics, computational chemistry, molecular evolution and membrane proteins.
Computational Biology and Bioinformatics

Masters of Science in Bioinformatics

Does the Masters of Science in Bioinformatics offer grants or financial assistance? Why or why not?

Grants, Teaching Assistantships and research fellowships are not available. This is a professional Masters Degree program.
Bioinformatics PhD

View tabbed content | View all content (for printing)

Career opportunities

In terms of career progression, the majority of our postgraduate students remain in research, while others move into related areas such as education, information technology, consultancy and management, helped by the comprehensive skills training they gain during their programme of study.
Bioinformatics MSc and PhD at Imperial College

General Information

For general details of Postgraduate study at Imperial see: http://www3.imperial.ac.uk/pgprospectus

Important note: Most funding bodies such UK Research Councils have eligibility requirements for PhD citizenship. You should check the web site of the funding body for the latest information.
UCD Bioinformatics and Systems Biology PhD Programme

PhD in Bioinformatics and Systems Biology

New high-throughput technologies in biology have opened up exciting opportunities for **numerate scientists** to work in advanced areas of **biological research** (see excellent video [here](http://bioinfo-casl.ucd.ie/PhD/)).

Our programme takes students from a variety of backgrounds (statistics, engineering, mathematics, computer science, biology, chemistry, physics) and gives them relevant courses during their PhD, tailored to their needs.

There are several projects available in different areas such as systems biology, statistics, cancer, signalling and infection.
I’ll study Bioinformatics...
ОБУЧЕНИЕ БИОИНФОРМАТИКЕ
Курсы по биоинформатике для молодых ученых

3-я Научная школа молодых ученых «Биоинформатика и системная биология»

Семинары | Биотехнологии, Биология

Россия, Новосибирск

Приглашаем вас принять участие в школе молодых ученых «Биоинформатика и системная биология», которая будет третьей, проведенной Институтом цитологии и генетики. Школа пройдет на базе Института цитологии и генетики с 22 по 25 сентября в Новосибирском Академгородке.

Целью научной школы является ознакомление студентов, аспирантов и молодых ученых с современными работами в области биоинформатики и системной биологии. Вашему вниманию будут представлены лекции ведущих российских и зарубежных ученых на сегодняшние достижения в области системной биологии, современным методам и конкретным примерам исследования. Ожидается, что некоторые из них будут связаны с генетическими сетями и процессами развития и многому другому. Программа школы также будет включать устные доклады участников школы. По результатам научных докладов молодых ученых лучшие работы будут номинированы.
Підручники з біоінформатики
Welcome

Welcome to the website for Bioinformatics and Functional Genomics (2nd edition) by Blackwell, 2009. This site features a complete bioinformatics teaching curriculum, powerpoints for an entire course taught at the Johns Hopkins School of Medicine, and website links organized by chapter in the new textbook, Bioinformatics and Functional Genomics. A separate resource (at the Wiley Higher Education site) is available for instructors, providing detailed solutions to problems.

Appendix 1: Glossary

This glossary is combined from five web-based glossaries: [1] the National Center for Biotechnology Information (entries marked NCBI), [2] the Oak Ridge National Laboratory (entries marked ORNL), [3] the talking glossary at the National Human Genome Research Institute, [4] the SMART database, and [5] the protein folds glossary from the Structural Classification of Proteins website (SCOP) (these entries are modified). Additional web-based glossaries are listed in a table at the end of this glossary.

The glossaries are on-line at:

A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z
Навчальні ресурси на біоінформаційних сайтах (NCBI)

NCBI Home
Site Map (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Training & Tutorials

| Databases | | | | | |
|-----------|---|---|---|---|
| All       | Databases | Downloads | Submissions | Tools | How To |

NCBI C++ Toolkit Manual
A comprehensive manual on the NCBI C++ toolkit, including its design and development framework, a C++ library reference, software examples and demos, FAQs and release notes. The manual is searchable online and can be downloaded as a series of PDF documents.

NCBI Education Page
Provides links to tutorials and training materials, including PowerPoint slides and print handouts.

NCBI Glossary
Part of the NCBI Handbook, this glossary contains descriptions of NCBI tools and acronyms, bioinformatics terms and data representation formats.

NCBI Handbook
An extensive collection of articles about NCBI databases and software. Designed for a novice user, each article presents a general overview of the resource and its design, along with tips for searching and using available analysis tools. All articles can be searched online and downloaded in PDF format; the handbook can be accessed through the NCBI Bookshelf.

NCBI Educational Resources

NCBI provides a variety of educational resources, including tutorials, problem sets, online tools, FAQs and documentation.

# Календарно-тематичний план заняття з курсу

«Основи біоінформатики» для студентів 3 курсу – 2015

<table>
<thead>
<tr>
<th>№</th>
<th>Дата</th>
<th>Тема заняття</th>
<th>Ауд.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>18.02</td>
<td>Застосування біоінформатики у біологічних дослідженнях. Біоінформаційні ресурси. Біологічна інформація. Використання біоінформаційних ресурсів в рамках курсових і дипломних робіт. Підходи до пошуку інформації в БД</td>
<td>3-15</td>
</tr>
<tr>
<td>2.</td>
<td>04.03</td>
<td>Пошук послідовностей. NCBI-Gene. Лінійна біоінформатика. Формати файлів. Вирівнювання. Пошук гомологів. BLAST-NCBI.</td>
<td>3-15</td>
</tr>
<tr>
<td>3.</td>
<td>18.03</td>
<td>Множинне вирівнювання послідовностей та його застосування. Cobalt. Clustal. Інструменти SwissProt (SIB). Філогенетичні дерева.</td>
<td>3-15</td>
</tr>
<tr>
<td>4.</td>
<td>01.04</td>
<td>Геномні браузери. Ensembl. UCSC.</td>
<td>3-15</td>
</tr>
<tr>
<td>5.</td>
<td>15.04</td>
<td>Структурна біоінформатика. PDB. Класифікація білкових структур.</td>
<td>3-15</td>
</tr>
<tr>
<td>6.</td>
<td>29.04</td>
<td>Структурна біоінформатика. Моделювання білкових структур.</td>
<td>3-15</td>
</tr>
<tr>
<td>7.</td>
<td>13.05</td>
<td>Ресурси системної біології. KEGG. Сигналінг. Enzyme. Генна онтологія. QuickGO. БД мікроРНК.</td>
<td>3-15</td>
</tr>
<tr>
<td>8.</td>
<td>27.05</td>
<td>Залікове заняття</td>
<td>3-15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Додатково – практичні заняття на кафедрі</td>
<td>(2-6)</td>
</tr>
</tbody>
</table>