Master Degree

Laurea Magistrale

BIOINFORMATICS FOR COMPUTATIONAL GENOMICS
The introduction of new technologies for genomic analysis, firstly the Next Generation Sequencing, has revolutionized the analysis methods in all areas of the biomolecular and biotechnological sciences. Indeed, these technologies allow generating multiple data types, such as the complete sequence of a genome, or the simultaneous measurement of expression levels of tens of thousands of genes, opening to researchers possibilities even unthinkable until few years ago.

The management and analysis of such data require the use of computational, statistical and of machine learning techniques able to overcome difficulties due to data dimension and heterogeneity. Thus, it has become needed the training of experts who own suitable knowledge both in computer science, statistics and computational areas for the data management, processing and quantitative analysis, and in biomolecular and genetic areas for the coherent and full biological interpretation of results.
High throughput sequencing cost per genome, 2001-2024

Cost per Human Genome

Moore’s Law

NIH
National Human Genome Research Institute

genome.gov/sequencingcosts
## Four domains of Big Data in 2025

<table>
<thead>
<tr>
<th>Data Phase</th>
<th>Astronomy</th>
<th>Twitter</th>
<th>YouTube</th>
<th>Genomics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition</td>
<td>25 zetta-bytes/year</td>
<td>0.5–15 billion tweets/year</td>
<td>500–900 million hours/year</td>
<td>1 zetta-bases/year</td>
</tr>
<tr>
<td>Storage</td>
<td>1 EB/year</td>
<td>1–17 PB/year</td>
<td>1–2 EB/year</td>
<td>2–40 EB/year</td>
</tr>
<tr>
<td>Analysis</td>
<td>In situ data reduction</td>
<td>Topic and sentiment mining</td>
<td>Limited requirements</td>
<td>Heterogeneous data and analysis</td>
</tr>
<tr>
<td></td>
<td>Real-time processing</td>
<td>Metadata analysis</td>
<td></td>
<td>Variant calling, ~2 trillion central processing unit (CPU) hours</td>
</tr>
<tr>
<td></td>
<td>Massive volumes</td>
<td></td>
<td></td>
<td>All-pairs genome alignments, ~10,000 trillion CPU hours</td>
</tr>
<tr>
<td>Distribution</td>
<td>Dedicated lines from antennae to server (600 TB/s)</td>
<td>Small units of distribution</td>
<td>Major component of modern user's bandwidth (10 MB/s)</td>
<td>Many small (10 MB/s) and fewer massive (10 TB/s) data movement</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pbio.1002195.t001

http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002195
High demand of computational scientists

With over 25 million human whole genome sequences expected by 2025, there is a huge need of training researchers/experts capable to:

• transform the enormous amount of biological information ("big data") into knowledge

• gain quantitative insight into the behavior of biological systems towards precision medicine

An Explosion Of Bioinformatics Careers

By Alaina G. Levine

Big data is pouring out of life sciences research, creating ample opportunities for scientists with computer science expertise.

• Big pharma, biotech, and software companies are clamoring to hire professionals with experience in bioinformatics and the identification, compilation, analysis, and visualization of huge amounts of biological and health care information.
Opportunities for computational scientists careers

"The opportunities are enormous for budding computer scientists, mathematicians, and engineers to meld with trainee biologists, neuroscientists, and clinicians in novel university courses constructed to focus on the truly 21st century phenomenon that is the data avalanche from large-scale biomedicine"

John Darrell Van Horn. Opinion: Big data biomedicine offers big higher education opportunities
Proc Natl Acad Sci U S A. 2016 Jun 7; 113(23): 6322–6324. Published online 2016 Jun 7. doi: 10.1073/pnas.1607582113
http://www.pnas.org/content/113/23/6322
Bioinformatics for Computational Genomics

- Bioinformatics and Systems Biology at Manchester University
- Bioinformatics and Computational Genomics at Queen’s University in Belfast
- Bioinformatics at Goethe University in Frankfurt
- Bioinformatics and Computational Biology at University of Idaho and at University of Texas.

At the Harvard School of Public Health has been recently presented the Master of Science in Computational Biology and Quantitative Genetics, aimed at providing rigorous training in data analysis in the current biomedical research contest, addressed to Information Engineering profiles who work in research centers, Universities, pharmacological industries and in the biotechnology sector.

The ETH in Zurich offers the Master in Computational Biology and Bioinformatics in cooperation with the University of Zurich and the University of Basel. The Imperial College of London offers the Master of Science in Bioinformatics and Theoretical Systems Biology.
Bioinformatics for Computational Genomics

In the last years, a few Master Degrees have been activated, based on different Bioinformatics aspects, in some Italian Universities, both in the class of the Biotechnology Degree (University of Milano, University of Bologna), and in the class of the Computer Science Degree (University of Verona).

Yet, the Degrees currently active are mainly focused on only a single aspect of Bioinformatics at a time, i.e., or the one for the use of analysis methods (Biotechnologies, e.g., the Master Degree in “Molecular Biotechnology and Bioinformatics” at University of Milano) or the one for the development of analysis algorithms and methods (Computer Science).
The only known examples that aim at the training of a real multidisciplinary profile, with both functional genomic and information processing and analysis expertise, are:

- The Degree in **Bioinformatics** (Degree class in Biology) at the University of Roma – Tor Vergata (where however extra credits/exams are given to students based on their 3-years Degree)

- The Master of Science in **Quantitative and Computational Biology** (QCB) at the University of Trento (where, however, students have to follow separate curricula based on their 3-years Degree, and they have to choose to deepen or the biological or the computer science aspects).
Bioinformatics for Computational Genomics

The Master Degree ("Laurea Magistrale") in “Bioinformatics for Computational Genomics” (LM-8 Biotecnologie Industriali) is a joint initiative of the Politecnico di Milano and the University of Milano that has been activated in the Academic Year 2019-2020, and will award a “Joint Degree” (awarded by both Universities).

Overall, it aims to form "genome data scientists", that is, graduates that will have an adequate knowledge about:

- the molecular basis of biological systems
- the structure and function of biological molecules and how they participate in cellular processes
- the technologies and platforms for the analysis of genomes
- the tools for bioinformatic and genomic analysis
- the statistical and computational methodologies for the analysis of biomolecular data
By attending the courses of the BCG degree you will learn about:

1) **Organization of information in the genome** and the molecular and cellular processes at the basis of **gene expression and regulation**

2) **Experimental methods** for studying genes and their function in different model species, both prokaryotic and eukaryotic

3) **Technologies** used in modern genomic and epigenomic research, like next-generation sequencing based assays

4) **Methods and protocols** of bioinformatic analysis in functional genomics studies

5) **Algorithmic, mathematical and statistical approaches** underlying bioinformatic and genomic analysis tools

6) **Data base technologies** for the storage and organization of the data

7) **Modelling and analysis techniques** employed in systems biology for the study of interactions in complex biological systems
Bioinformatics for Computational Genomics
Interdisciplinary courses

• Will emphasize quantitative and computational aspects, with a focus on modeling, analyzing and understanding biological systems and phenomena

• Will enable students to face essential questions at the interface between fundamental research and clinical and pre-clinical areas

• Being offered by two complementary universities, it will provide to its students a truly multidisciplinary context:
  • In the first semester, based on their first-level degree, students will follow two different tracks to integrate their preparation. Then, they will follow a unique track, with courses both on computational and biotechnological topics
  • Will produce four professional profiles: biological data and systems analysts, computational biologists, bioinformaticians, and biotechnologists.
Professional profiles at outcome

**BIOLOGICAL DATA AND SYSTEMS ANALYST**
- Skills in biostatistics for experimental design and advanced analysis of clinical and molecular data
- Curation of high-dimensional data for the representation of biological knowledge by integrated systems and biological networks
- Analysis of complex biological data derived from whole-cell measurements

**BIOINFORMATICIAN**
- Development of software and computational tools for biological data analysis
- Skills in complex databases, web resources and in cluster- or cloud-based computational solutions
- Building and maintaining workflows and pipelines for raw biomolecular data analysis and integration

**COMPUTATIONAL BIOLOGIST**
- Development of analytical methods for large-scale data (genomics, proteomics, metabolomics) and molecular interaction models
- Identification, development and test of computational methods for the discovery of active ligands and for drug development
  - Design of in silico experiments to test data-driven hypotheses and interpretation of result

**BIOTECHNOLOGIST**
- Skills in experimental techniques and instrumentation and deep understanding of cellular and molecular processes
- Skills in dedicated software for laboratory equipment and raw data management
- Effective interaction both with laboratory personnel and analysts and ability to integrate workflows
### Bioinformatics for Computational Genomics

#### Courses

<table>
<thead>
<tr>
<th>UniMI</th>
<th>CFU</th>
<th>SSD</th>
<th>PoliMI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st year - 1st semester</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetics</td>
<td>12</td>
<td>BIO/18</td>
<td>Programming</td>
</tr>
<tr>
<td>Cellular and Molecular Biology</td>
<td></td>
<td>BIO/11</td>
<td>Data Bases</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>6</td>
<td>BIO/10</td>
<td>Statistics</td>
</tr>
<tr>
<td>Organic Chemistry</td>
<td>6</td>
<td>CHIM/06</td>
<td>Bioinformatics and Computational Biology</td>
</tr>
<tr>
<td><strong>1st year - 2nd semester</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genomics</td>
<td>12</td>
<td>BIO/11</td>
<td>Scientific Programming</td>
</tr>
<tr>
<td>Transcriptomics</td>
<td></td>
<td>BIO/11</td>
<td>Biostatistics</td>
</tr>
<tr>
<td><strong>2nd year - 1st semester</strong></td>
<td></td>
<td></td>
<td>Machine Learning</td>
</tr>
<tr>
<td>Advanced Genomics</td>
<td>12</td>
<td>BIO/19</td>
<td>Systems Biology and Network Analysis</td>
</tr>
<tr>
<td>Epigenomics</td>
<td></td>
<td>BIO/18</td>
<td>Genomic Big Data Management and Computing</td>
</tr>
<tr>
<td>Structural Chemistry</td>
<td>6</td>
<td>CHIM/06</td>
<td>Interdisciplinary Project</td>
</tr>
<tr>
<td>Neurogenomics and Brain Disease Modelling</td>
<td>6</td>
<td>BIO/11</td>
<td></td>
</tr>
<tr>
<td><strong>Additional requirements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistics Skills &amp; other</td>
<td>3</td>
<td></td>
<td>Alignment Plan 1 (graduated in computer science, mathematics, or similar)</td>
</tr>
<tr>
<td>Training + Final Test</td>
<td>21</td>
<td></td>
<td>Alignment Plan 2 (graduated in biology/biotechnology, or similar)</td>
</tr>
<tr>
<td><strong>Free choice courses</strong></td>
<td></td>
<td></td>
<td>All students</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Shared with Ing. Inf. &amp; Ing. BioMedMaster Degree</td>
</tr>
</tbody>
</table>

Classes both at the "Città Study Campus" of the University of Milano (UniMI), and at the "Leonardo Campus" of the Politecnico of Milano (PoliMI). The two campuses are within walking distance from one another.
The program includes, as a fundamental step in the formation of the students, an internship in research laboratories either at the University of Milan, the Politecnico, or in other Italian or foreign research institutes of at least six months.

You will work on a project on the development or application of bioinformatic tools and pipelines for the analysis or the organization of genomic data, either in basic or applied research.

The research experience of the internship and its results will be described in a final written dissertation, to be discussed in front of a thesis committee.
Bioinformatics for Computational Genomics

Admission

The admission prerequisite is a Bachelor Degree ("Laurea Triennale") providing an adequate knowledge of either:

1. genetics, molecular biology and biochemistry, or
2. computer science, information engineering and mathematics

- Knowledge of the English language is required (B2 level; English proficiency, assessed during the interview)

Students, also who still have not completed their Bachelor Degree, can apply at [https://bcg.cdl.unimi.it/en/enrolment](https://bcg.cdl.unimi.it/en/enrolment) (from March 14th to June 27th (May 30th for non EU), at 2.00 PM), if they plan to graduate before October 31st, 2024

For the academic year 2024-2025, a maximum of 50 students (+ 10 non-EU) will be admitted, based on their background knowledge and a personal interview (on July 4th, 2024).
1) They have completed a Bachelor Degree program (Laurea Triennale) in **one of the following classes:**

- Biotechnology (class L2)
- Biology (class L13)
- Agriculture and Food Sciences (class L26)
- Pharmacological Sciences (class L29)

**and acquired at least 30 CFU in biological areas (SSD BIO) with at least 18 CFU in Genetics (BIO/18), Molecular Biology (BIO/11) and Biochemistry (BIO/10)**
2) They have completed a Bachelor Degree program (Laurea Triennale) in one of the following classes:
- Information Engineering (class L8)
- Computer Sciences (class L31)
- Physics (class L30)
- Mathematics (class L35)
and acquired at least 30 CFU in the areas of computer science, information engineering, biomedical engineering, mathematics and/or statistics (SSD INF/01, ING-INF/05, ING-INF/06, MAT/01-09, and/or SECS-S/01), with at least 6 CFU in mathematics (MAT/01-09) and at least 12 CFU in one or more of the following areas: computer science (INF/01), information engineering (ING-INF/05), biomedical engineering (ING-INF/06), statistics (SECS-S/01)
The same criteria are applied to candidates in possession of foreign University Degrees deemed to be suitable by the Teachers Council, in which it is possible to clearly identify disciplines and number of credits acquired for each discipline.

If this is not possible, the documents certifying the career of the candidates will be examined in detail by the Teachers Council in order to assess if their background complies with the previous requirements.
Places available: 50 + 10 reserved for non-EU citizens

Call for applications
Please refer to the call for admission test dates and contents, and how to register.

Application for admission: from 14/03/2024 to 27/06/2024
Admission test date and venue: 04/07/2024 - da remoto
Application for matriculation: from 09/07/2024 to 18/07/2024

Read the Call

Programme description and courses list

Programme description (Manifesto)
Educational Plan

Official documents
Programme teaching regulations - F4B
Bioinformatics for Computational Genomics
Selection and ranking

Students satisfying the prerequisites will be admitted to an interview with members of Teaching Board for the assessment of their scientific background knowledge.

The adequate personal preparation of the candidates and expertise in their bachelor degree background topics, their ability to communicate in English and their motivation are decisive elements for the admission.

For the Academic Year 2024-2025, interviews for Italian students and foreign students residing in Italy will take place online only, on July 4th 2024, h.9.30.

Foreign applicants not resident in Italy and who achieved their Bachelor degree abroad will be evaluated based on their CV only.
Bioinformatics for Computational Genomics

Enrolling after admission

Once the results have been published (9/7/2024, after 14:00), students who occupy the first 50 positions in the merit ranking must complete their enrolment between 9/7/2024, 14:00 and 18/7/2024, 14:00

Subsequently, if after the initial deadline the first 50 available positions are not saturated, enrollment will be possible between 19/7/2024, 14:00 and 25/7/2024, 14:00 for students with lower merit ranking, provided they have achieved at least 60/100 in the admission test

Enrollment of undergraduates will be validated only if they obtain their Bachelor Degree (Laurea Triennale for Italian students) by 31 October 2024
Bioinformatics for Computational Genomics
Second Degree in Biomedical or Computer Science Engineering

Students with a Bachelor Degree (Laurea Triennale) in *Biomedical Engineering* or *Information Engineering* who obtained the Master Degree in *Bioinformatics for Computational Genomics* will also have the opportunity of enrolling as second degree in the Master Degree program in, respectively, *Biomedical Engineering* or *Computer Science and Engineering* of Politecnico di Milano with the automatic transfer in their career of several of the credits acquired in the *Bioinformatics for Computational Genomics* program, obtaining a very significant reduction of the number of exams to be taken to obtain the additional Master Degree in *Biomedical Engineering* (only 40 CFU, within a pre-approved study plan) or *Computer Science and Engineering* (only 35 CFU compulsory and 10 CFU at student’s choice)
Questions?