PA Bioinformatique X23 2025/26

Sebastian.Will@polytechnique.edu
Sarah.Berkemer@polytechnique.edu

AMIBio

Apr 2025

PA (Programme d'Approf)—A fundamental change

The PA is the **first year of your Master**—not *only* a continuation of 1A/2A

- specialization
- specific scientific field
- building your profile
- acquire concrete, specific skills

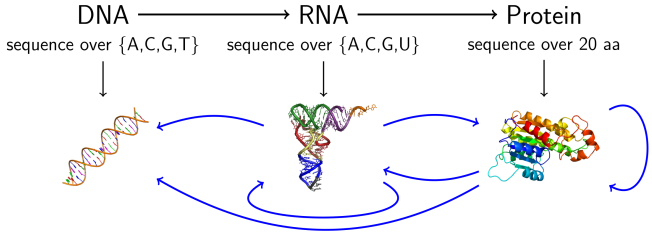
Start to think about

- 4A, continuation of the Master program
- PhD thesis
- start into your professional career

Find essential info in the livret PA and at 

`https://www.enseignement.polytechnique.fr/bioinformatique`

Bioinformatics: computational molecular biology and more



- modern Biology, Medicine, Neuroscience... requires Computer Science to handle ***Massive Amounts of Data***
  - *high-throughput* sequencing analysis: genomics, transcriptomics, ...
  - *comparative* genomics, phylogenomics
  - classification and *annotation* of proteins and RNAs
  - *structure modelling and design* of new bio-molecules; "drug design"
  - ...
- new possibilities in ***Biology***, ***Pharmacology***, ***Medicine***, ...
- challenges in ***Modeling***, ***Algorithmics***, ***Combinatorics***, ***Machine Learning***...

Bioinformatics: computational molecular biology and more

- modern Biology, Medicine, Neuroscience... requires Computer Science to handle ***Massive Amounts of Data***
  - *high-throughput* sequencing analysis: genomics, transcriptomics, ...
  - *comparative* genomics, phylogenomics
  - classification and *annotation* of proteins and RNAs
  - *structure modelling and design* of new bio-molecules; "drug design"
  - ...
- new possibilities in ***Biology***, ***Pharmacology***, ***Medicine***, ...
- challenges in ***Modeling***, ***Algorithmics***, ***Combinatorics***, ***Machine Learning***...

Special Challenge: multi-inter-disciplinary — CS, biology, ...

Objective: dual competence - interact with biologists, medics, computer sciencists, ... in academia and industry

Many flavors of Bioinformatics

= analysing - modeling - structuring - explaining - predicting - handling ...
biologically relevant information

= (Medicine +) Biology + Computer Science (+ Math, Statistics)

= knowledge of biology and sophisticated techniques, e.g.

- data search, index structure
- combinatoric optimization
- classification, machine learning
- visualization, image analysis

Questions/challenges for Bioinformatics

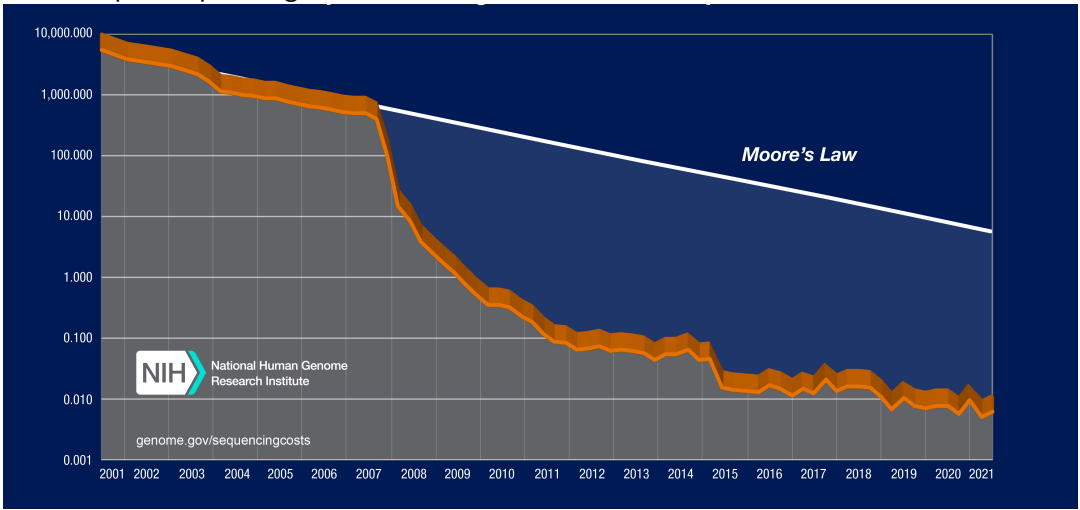
- sequences, genomics, gene identification, evolution,
- pangenomics, ..., personalized (genomic) medicine,
- structures (proteins, RNAs, complexes, ...),
- design/engineering problems: design of structures, functions, drugs,
- cell functions (metabolism, mitosis, ...),
- bio-molecule interactions (regulation of genes, regulatory networks...),
- statistics: should we be surprised to...?

Need for grammatical, combinatorial, probabilistic models
$\longrightarrow$ *algorithms:* efficient exact algorithms, approximations, heuristic strategies, ML

Challenge of big data

data volume increases much faster than computing power

For example, sequencing cost:



Challenge of big data

data volume increases much faster than computing power

- search
- indexing
- filtering
- statistics
- extraction of textual information
- building and using ontologies
- high dimensions, large graphs, ...

and also engineering issues (hardware and software), systems architecture

4th year after the Bioinfo PA

- M2 Bioinformatics in France, e.g.:
  - AMI2B de Paris-Saclay,
  - parcours BIM de UPMC,
  - Bordeaux, Aix-Marseille, Toulouse, ...
- abroad, MSc in 12 month (system 4+1), e.g.:
  - Cambridge: MPhil in Computational Biology,
  - Imperial College London :
  - MSc in Bioinformatics and Theoretical Systems Biology,
  - Edinburgh: MSc in Bioinformatics,
- McGill, Montreal: MSc in Computer Science / Bioinformatics,
- Europe, complete MSc: 2 years (system 3+2), e.g.:
  - ETH Zürich: Master in Computational Biology and Bioinformatics,
  - EPFL (catégorie 1), Master in Life Sciences Engineering
  - Copenhagen (UofC and DTU),
  - Leipzig, Freiburg, Vienna... (contact me)
- Ecoles: Agro, Mines, ...

Diverse bionformatics-related master programs, e.g.

- Master of BioEngineering
  - e.g.: EPFL, Stanford, Berkeley
- Biology and Data Science / Biostatistics, e.g.:
  - Harvard: MSc of Science in Computational Biology and Quantitative Genetics ("Big Data", stats), MSc in Health Data Science, ...
  - Columbia,
  - Yale, ...
- Neurosciences, e.g.:
  - Oxford: MSc in Neuroscience,
  - EPFL: Master in Life Sciences Engineering / Neurosciences and Neuroengineering,

Opportunities after 4A

PhD thesis (recommended) $\rightarrow$ appropriate modules in master (or PhD track)

- Industry :
  - pharmaceutical (drug design),
  - agro-food (yields, climate, palatability, etc.),
  - biotechnologies (fuels, materials, etc.),
  - environment,
  - ...
  - IT (medical imaging, care systems, support, etc.) e.g. Dassault Systemes, IBM, GE, Siemens, ...
- Large institutes, e.g. : Curie, Pasteur, INRAE, INSERM, ..., EMBL-EBI, SIB, NCBI, ...
- Academia

Requirements for the PA

Prerequisites from BIO + INF:

- 1 Biology course in 2A: ..., BIO452 (recommended), ...
- 2 computer courses in 2A (excluding modal) ..., INF421, ..., INF442, ... (recommended)
- 1 CS modal or 1 project integrated into a course (some programming experience)
- MAP433 (statistics) recommended
- **also check course-specific prerequisites**

see PA homepage; discuss special cases with me

Rules for program choices

- Each period: 3 classes + 1 EA (or long project)
  - total 8 including at least 3 in biology and 3 in computer science (several “projects”)
  - each period: possibility of a course outside the program
    generally ok, if course BIO or INF
    others: must be motivated
  - *subject to schedule compatibility*
- Long project INF511
  recommended: opportunity to inter-disciplinary transfer...
  INF511 replaces 2 EAs in P1&2
- 3A internship in Biology (BIO591), Computer Science (INF591), Data Science (INF592); final choice in the fall

also: non-standard options → adaptation from the PA Bio or the PA Info, PA SDE...

PA program: classes

Term 1

3 classes from

- BIO551 Immunologie et agents infectieux
- BIO553 Biotechnologies pour la médecine et l'agriculture
- BIO556 Genomes, diversity, environment and human health
- BIO557 Neurosciences
- INF550 Algorithmique avancée
- INF552 Data Visualization
- INF555 Constraint-based Modeling and Algorithms for Decision Making
- INF556 Topological data analysis

Term 2

obligatory

- INF589 Computational analysis of high-throughput sequencing data

2 classes from

- BIO562 Biologie des systèmes moléculaires
- BIO563 Epigénétique et ARN non-codants
- INF580 Large scale mathematical optimization
- INF581 (Advanced) Topics in Artificial Intelligence
- MAP566 Statistics in action
- MAP569 Machine learning II

PA program: projects and internship

Term 1

1 EA (or long project)

BIO571A	Travaux expérimentaux de génie génétique
BIO571B	Travaux expérimentaux en imagerie quantitative
INF554	[EA] Machine learning I
INF573	[EA] Image Analysis and Computer Vision

Terms 1&2: long projects

INF511	Projet de Bioinformatique

Term 2

1 EA (or long project)

BIO583	Sciences des données en imagerie biologique
BIO/INF588	Projet en bioinformatique

Term 3

Stage / internship

BIO591	Biologie et Écologie
INF591	Informatique
INF592	Data Science

INF511 Projects

Examples from past years:

- Models for protein coevolution
- Graph neural networks for clustering of RNA structures
- A Divide and Conquer Approach for RNA Design
- Prediction of nucleotides at protein/RNA interfaces
- Classification of protein structures
- Implementation of a stochastic simulation algorithm for metabolic networks
- Classification of electroencephalograms and detection of epileptic seizures
- Executable mathematical model of a single red blood cell
- A prediction algorithm for interactions of multiple triplet-repeat RNAs
- Non-invasive prenatal diagnosis of monogenic diseases
- Benchmarking of single-cell RNA-seq
- Alignment of brain imaging data

start planning before summer

The (not so) far future: 3A internship and 4A/M2

Start thinking about

- your internship:

  BIO591 Biology and Ecology internship (Yves Mechulam)
  INF591 Computer Science internship (Benjamin Werner)
  INF592 Data Science internship (Ioana Manolescu, Jesse Read)

- your ideas for 4A/M2

Start to discuss with researchers, teachers, other students...

Stage de recherche - some examples

- SFU, Vancouver : Computational tools for human pathogens outbreak monitoring using whole-genome sequencing publication
- Paris : New technology for neuropathies and data mining : Application to Cystic Fibrosis and Prediabetes screening (prix de stage)
- MIT : Machine Learning applied to the Amyotrophic Lateral Sclerosis
- IMAG : Evaluation of distance between Thresholded Boolean Automaton Networks (prix de stage), publications
- University of Sherbrooke (distance, phylogénie, cancers)
- University of Vienna: Advanced boundary prediction of non-coding RNAs
- EBI/EMBL : Analysis of Single-Cell RNA Sequencing data from human pancreas
- Mc Gill : Leveraging Affinity Information to improve Molecular Generative Models (prix de stage)
- GMT Science, Paris: AI Methods to identify microbiome-based signatures
- Doctolib SAS, Paris: Intelligent agents for medical appointments

Timeline

- start of **inscription** (“now”)
  - PA choice, motivation letter, course choice
  - individual discussions
- end of inscription and validation (“too soon”)

- Back-to-school meeting (September)
- Individual discussion (change of courses, project, internship, 4A)
- Deadlines
  - November: limit category 1
  - December: category 2 and 3 limit, internship limit
  - January: last applications abroad

À bientôt ...

You are welcome to discuss and ask for more details ...

Sebastian Will
sebastian.will@polytechnique.edu
https://www.lix.polytechnique.fr/~will

Sarah Berkemer
sarah.berkemer@polytechnique.edu

Bionformatics team
at DIX/LIX
https://www.lix.
polytechnique.fr/amibio

https://synapses.polytechnique.fr

https://www.enseignement.
polytechnique.fr/
bioinformatique