

PA Bioinformatique X21

Sebastian.Will@polytechnique.edu

Sarah.Berkemer@polytechnique.edu



September 2023

largely based on slides of Philippe Chassignet

Rules of the PA

- Each period: 3 classes + 1 EA (or long project)
 - total 8
 - ≥ 3 BIO, ≥ 3 INF (BIO/INF588, either bio or inf)
 - INF589 mandatory

Rules of the PA

- Each period: 3 classes + 1 EA (or long project)
 - total 8
 - ≥ 3 BIO, ≥ 3 INF (BIO/INF588, either bio or inf)
 - INF589 mandatory
- Long project (BIO511, BIO512, BIO572, INF511), replaces 2 EAs

starts now!

Rules of the PA

- Each period: 3 classes + 1 EA (or long project)
 - total 8
 - ≥ 3 BIO, ≥ 3 INF (BIO/INF588, either bio or inf)
 - INF589 mandatory
- Long project (BIO511, BIO512, BIO572, INF511), replaces 2 EAs **starts now!**
- **OK?** (at least P1) **you need our ok for changes**

Rules of the PA

- Each period: 3 classes + 1 EA (or long project)
 - total 8
 - ≥ 3 BIO, ≥ 3 INF (BIO/INF588, either bio or inf)
 - INF589 mandatory
- Long project (BIO511, BIO512, BIO572, INF511), replaces 2 EAs **starts now!**
- **OK?** (at least P1) **you need our ok for changes**
- **Don't** underestimate the *continuous* workload of the PA
(labs, challenges, projects, reading)
 - the end of P1 is dense (4A applications + seek for an internship)
 - the end of P2 is very tight (exams and defenses)

Rules of the PA

- Each period: 3 classes + 1 EA (or long project)
 - total 8
 - ≥ 3 BIO, ≥ 3 INF (BIO/INF588, either bio or inf)
 - INF589 mandatory
- Long project (BIO511, BIO512, BIO572, INF511), replaces 2 EAs **starts now!**
- **OK?** (at least P1) **you need our ok for changes**
- **Don't** underestimate the *continuous* workload of the PA
(labs, challenges, projects, reading)
 - the end of P1 is dense (4A applications + seek for an internship)
 - the end of P2 is very tight (exams and defenses)
- **3A internship** in Biology (BIO591) or Computer Science (INF591)
 - research internship (!)
 - to be evaluated by a single department BIO or INF
 - must be validated by the department \rightsquigarrow **choose attachment to BIO or INF**

stage 3A

- BIO 591, register with dept. BIO ↔ interview with Yves Mechulam
- INF 591 / INF 592 (more at PA Info; 17h amphi Gay-Lussac)
- in all cases :
 - "fiche de stage" (host institution, scientific tutor, subject)
 - + signed by the dept.
 - + a designated referent teacher \neq tutor
 - finally → SOIE for the "convention de stage"

stage 3A

- BIO 591, register with dept. BIO ↪ interview with Yves Mechulam
- INF 591 / INF 592 (more at PA Info; 17h amphi Gay-Lussac)
- in all cases :
 - "fiche de stage" (host institution, scientific tutor, subject)
 - + signed by the dept.
 - + a designated referent teacher \neq tutor
 - finally \rightarrow SOIE for the "convention de stage"
- Contacts:
 - BIO591** Biology and Ecology internship
yves.mechulam@polytechnique.edu
 - INF591** Computer Science internship
olivier.bournez@polytechnique.edu
 - INF592** Data Science internship
ioana.manolescu@inria.fr, steve.oudot@inria.fr

Research internship INF591, examples

- Human pathogens outbreak monitoring using whole-genome sequencing, SFU, Vancouver
- Neuropathies and data mining : application to Cystic Fibrosis and Prediabetes screening, Impeto Medical, Paris
- Mining sparse statistical learning models from metagenomics data, ICAN, Paris
- Vision : cartes de fréquences spatiales et de sélectivités à la direction, CIRB, Collège de France, Paris
- Deep learning for gigapixel histopathology image analysis, JFLI, Tokio

Research internship INF591, examples, ctd

- Identifying structure within clinical and genomic data for ALS, Biological Engineering, MIT
- Developing a learning framework for end-to-end docking prediction, Computer Science, McGill, Montreal
- Alignment of short query sequences against large probabilistic genomes, Computer Science, McGill, Montreal
- Recherche de bassins d'attraction dans des réseaux biologiques, LS2N, Nantes
- Evaluation of distance between Thresholded Boolean Automaton Networks, TIMC, Grenoble
- Application of persistence theory to the clustering of neurons, Blue Brain Project, EPFL, Lausanne

Research internship INF591, examples, ctd

- Predicting effects of genomic variants in Amyotrophic Lateral Sclerosis, Biological Engineering, MIT
- Analysis of single-cell RNA-seq data from human pancreas, EMBL-EBI (Hinxton, UK)
- Prediction of TAD (Topologically Associating Domains) conformation across evolution rearrangement scenarios, McGill Center for Bioinformatics
- Leveraging Affinity information to improve molecular generative models, McGill Center for Bioinformatics
- Sophisticated detection of RNA consensus structure domain boundaries, TBI, U Vienna
- Algorithms for edit distances of tutor genomes by duplications and deletions, U Sherbrooke

Research internship INF591, examples, ctd

- Enlightening the analysis of protein interfaces with Multiple Interface String Alignment : application to the spikes of coronaviruses, INRIA Sophia Antipolis
- Additional study of pan-cancer Computational Histopathology reveals mutation, tumor composition and prognosis, EMBL-EBI (Hinxton, UK)
- A 3-dimensional study of RNA-DNA interactions using data provided by the RADICL-seq method, McGill Computational genomics lab
- Discovery of functional RNA motifs, McGill Computational genomics lab, Machine Learning for Molecular Biology, University of Edinburgh
- Sampled biomolecular energy landscapes : towards a stratification scheme, INRIA Sophia Antipolis
- Enhancing molecular docking using Graph Neural Networks, Aqemia (Paris)

4A, examples (Bioinformatics); different 'paths'

- M2 in Bioinformatics in France, ex :
 - new master at IP-Paris ?
 - BIM track at Sorbonne U.,
 - AMI2B at Paris-Saclay,
 - (MVA, very selective),
- MSc in 12 months (4+1 system), ex. :
 - Cambridge : MPhil in Computational Biology,
 - Imperial College London :
 - MSc in Bioinformatics and Theoretical Systems Biology,
 - Edinburgh : MSc in Bioinformatics,
- McGill : MSc in Computer Science / Bioinformatics, 12 months ...
- Europe, full MSc 2 years (3+2 system), ex. :
 - ETH Zürich : Master in Computational Biology and Bioinformatics,
 - EPFL : Master in Life Sciences Engineering / Computational Biology,
 - Copenhagen (U. of C. and DTU),
 - Germany (Freiburg, Leipzig,...) , Vienna, ...
 - french "écoles", ex. Agro, Mines, ...

4A, Masters 'outside' of bioinformatics

- Health and Data Science (Biostatistics), ex. :
 - Harvard : MSc of Science in Computational Biology and Quantitative Genetics,
 - MSc in Health Data Science, ...
 - Columbia, ...
- Neurosciences, ex. :
 - Oxford : MSc in Neuroscience,
 - EPFL : Master in Life Sciences Engineering / Neurosciences and Neuroengineering,

What next?

PhD (recommended) \rightsquigarrow plan ahead when choosing & during the master (PhD track !)

Then, many options

- industry :
 - pharmaceutical (drug devel, applied medicine),
 - agriculture/food (yields, climate, taste, ...),
 - biotechnologies (fuels, materials, ...),
 - environment,
 - computer science (imaging, ...)
ex. Dassault Systèmes, IBM, GE, Siemens, ...
 - ...
- major institutes, ex. :
Curie, Pasteur, INRA, INSERM, ...,
EMBL-EBI, SIB, NCBI, ...
- academic career

See you around...

You are welcome to discuss details / get more info ...



Sebastian Will
sebastian.will@polytechnique.edu



Bionformatics team
at DIX/LIX



Sarah Berkemer
sarah.berkemer@polytechnique.edu

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

<https://synapses.polytechnique.fr>