

# PA de Bioinformatique promotion X18

Philippe.Chassignet@polytechnique.edu

23 avril 2020

lien :

[http://www.enseignement.polytechnique.fr/profs/informatique/Philippe.Chassignet/PA\\_bioinformatique/](http://www.enseignement.polytechnique.fr/profs/informatique/Philippe.Chassignet/PA_bioinformatique/)

<https://synapses.polytechnique.fr/catalogue/2020-2021/parcours/...> (pas encore là)

# Programme d'Approfondissement

- ce n'est pas (que) une suite de la 2A !
- c'est la première année d'un programme master,
- une thématique de master :
  - = un choix de spécialisation,
  - = un champ disciplinaire,
  - = un profil lisible,
  - = de réelles compétences,
- c'est un choix cohérent de cours, projet et stage,
- ce sont des contacts ... d'autres contacts ...

Il faut déjà penser à la suite :

- (4A) complétion d'un master
- + thèse ?
- début de carrière



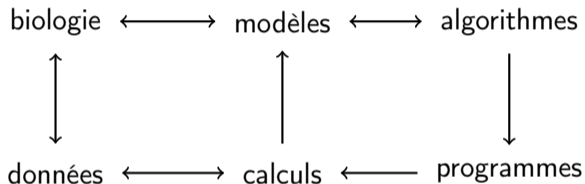




# Bioinformatique

= analyser, modéliser, structurer, expliquer ou prédire et traiter  
les informations relevant de la biologie

= ( Médecine + ) Biologie + Informatique ( + Mathématiques, Statistiques )



= des connaissances en biologie + des techniques sophistiquées, ex. :

- indexation, exploration combinatoire,
- fouille de données,
- classification, apprentissage automatique,
- visualisation, analyse d'image

# Bioinformatique

enjeux :

- séquences, génomique, identification de gènes, évolution,
- pangénomique, ..., médecine (génomique) personnalisée,
- structures (protéines, ARNs, complexes, ...),
- problème inverse : synthèse de formes, de fonctions, de médicaments,
- mécanismes de la cellule (métabolisme, mitose, ...),
- interactions (régulation des gènes, ...),
- statistiques : faut-il s'étonner de ... ?

besoin de modèles grammaticaux, combinatoires, probabilistes

↪ algorithmes : approximations, heuristiques, stratégies, ...

↪ apprentissage automatique

# Données massives

**le volume des données produites augmente beaucoup plus vite que les caractéristiques des ordinateurs**

problématique des données massives : produire de la connaissance

⇒ à ne pas réduire à *l'apprentissage statistique*

- fouille
- indexation
- filtrage
- statistiques
- extraction d'informations textuelles
- construction et utilisation d'ontologies
- hautes dimensions, grands graphes, ...

↪ et aussi des problèmes d'ingénierie (matérielle et logicielle), architecture des systèmes



## Après le PA, exemples

- M2 Bioinformatique en France, ex. :
  - AMI2B de Paris-Saclay,
  - parcours BIM de UPMC,
  - Bordeaux, Aix-Marseille, Toulouse, ...
- à l'étranger, MSc en 12 mois (système 4+1), 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*,
- Europe, MSc complet  $\approx$  2 ans (système 3+2), ex. :
  - ETH Zürich :  
*Master in Computational Biology and Bioinformatics*,
  - EPFL (catégorie 1), *Master in Life Sciences Engineering*
  - Copenhagen (U. of C. et DTU),
  - Allemagne, ...
- écoles, ex. Agro, Mines, ...

## Suite, exemples

- *Master of BioEngineering*
  - ex. : EPFL, Stanford, Berkeley
- Biologie et Science des Données, anc. Biostatistique, ex. :
  - Harvard : *MSc of Science in Computational Biology and Quantitative Genetics* (“Big Data”, stats), *MSc in Health Data Science*, ...
  - Columbia,
  - Yale, ...
- Neurosciences, ex. :
  - Oxford : *MSc in Neuroscience*,
  - EPFL : *Master in Life Sciences Engineering / Neurosciences and Neuroengineering*,

# Débouchés

Thèse (conseillée)  $\rightsquigarrow$  modules appropriés en master (ou PhD track)

- industrie :
  - pharmaceutique (médicament sur mesure),
  - agro-alimentaire (rendements, climat, appétence, ...),
  - biotechnologies (carburants, matériaux, ...),
  - environnement,
  - ...
  - informatique (imagerie médicale, systèmes de soin, support, ...) ex. Dassault Systèmes, IBM, GE, Siemens, ...
- grands instituts, ex. :  
Curie, Pasteur, INRA, INSERM, ...,  
EMBL-EBI, SIB, NCBI, ...
- académique

# Prérequis du PA

prérequis pour BIO + prérequis pour INF :

- 1 cours de Biologie en 2A : ..., BIO452 (conseillé), ...
- 2 cours d'Informatique en 2A (hors modal)  
..., INF421, ..., INF442, ... (conseillés)
- 1 modal d'Informatique ou 1 projet intégré à un cours,
- MAP433 conseillé,
- **vérifier les prérequis par cours**

profil 2A moyen (entrants dans le PA) :

	X15	X16	X17
BIO	3.2	2.7	2.9
INF	3.6	3.4	2.3
MAP	0.8	0.9	1.0
	<hr/>	<hr/>	<hr/>
	7.6	6.9	6.1
	/ 11	/ 11	/ 8

# Règles du PA

- chaque période : 3 cours + 1 EA (ou projet)
  - total 8
  - dont au moins 3 de biologie et 3 d'informatique  
( = plusieurs "projets" ou études d'articles )
  - chaque période : possibilité d'un cours hors programme  
ok si cours BIO ou INF  
autre  $\rightsquigarrow$  à motiver (sur i3a)
  - **sous réserve de compatibilité d'emploi du temps**
- projet long (BIO511, BIO512, BIO572, INF511)  
**recommandé**  $\rightsquigarrow$  utilisation "transversale" des compétences  
remplace 2 EA ou optionnel (éq. 2 modules sup.)
- stage de 3A en Biologie (BIO591) ou en Informatique (INF591)  
idem "transversale", choix définitif à l'automne

choix plus extrême  $\rightsquigarrow$  aménagement du PA de Bio ou du PA d'Info, PA SDE, ...

# Période 1

- 3 cours au choix parmi 8 :
  - BIO551 – Immunologie et agents infectieux.
  - BIO553 – Biotechnologies pour la médecine et l'agriculture. (nouveau)
  - BIO556 – Genomes, diversity, environment and human health.
  - BIO557 – Neurosciences.
  - INF550 – Algorithmique avancée.
  - INF552 – Data Visualization.
  - INF555 – Constraint-based Modeling and Algorithms for Decision Making Problems.
  - INF556 – Topological data analysis.
- 1 EA au choix parmi 5 (ou projet long) :
  - BIO571A – Travaux expérimentaux de génie génétique.
  - BIO571B – Travaux expérimentaux en imagerie quantitative.
  - INF554 – [EA] Machine learning I.
  - INF573 – Analyse d'images et vision par ordinateur
  - MAP573 – Data Analysis and Unsupervised Learning

## Période 2

- 3 cours au choix parmi 7 :
  - BIO562 – Biologie des systèmes moléculaires.
  - BIO563 – Epigénétique et ARN non-codants
  - BIO566 – Biologie computationnelle.
  - INF580 – Large scale mathematical optimization.
  - INF581 – (Advanced) Topics in Artificial Intelligence.
  - MAP566 – Statistics in action.
  - MAP569 – Machine learning II.
- 1 EA au choix parmi 2 (ou projet long) :
  - BIO583 - Sciences des données en imagerie biologique. (nouveau X17)
  - BIO/INF588 - Projet en bioinformatique

# Projets INF511

exemples des années passées :

- A Divide and Conquer Approach for RNA Design
- Prédiction des nucléotides aux interfaces protéine/ARN
- Classification de structures protéiques
- Implementation of Stochastic Simulation Algorithm, (for metabolic network)
- Classification d'électroencéphalogrammes et détection de crises d'épilepsie
- Executable mathematical model of a single red blood cell
- Monitoring of Anesthesia
- Diagnostic Prénatal Non Invasif de Maladies Monogéniques

commencer à y penser avant l'été

↪ on doit se voir pour en parler

**voir aussi BIO511, BIO512 ou BIO572**



# Stage de recherche

- BIO591 ou INF591 (pas INF592)
- il y a des stages en *Science des Données appliquée à la Biologie* dans INF591
- présentations en septembre
- catalogues ...
- contacts directs (mieux) :
  - stage **de recherche**
  - discuter avec enseignants ou chercheurs

## Stage de recherche INF591, exemples

- X14 :
  - 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
- X15 :
  - Mc Gill : *Adressage des protéines*
  - Mc Gill : *Alignment of short query sequences against large probabilistic genomes*
  - MIT : *Machine Learning applied to the Amyotrophic Lateral Sclerosis*
  - IMAG : *Evaluation of distance between Thresholded Boolean Automaton Networks*  
~→ prix de stage, publications
- X16 :
  - U. Sherbrooke (distance, phylogénie, cancers), U.Wien (algo ARN),
  - Mc Gill (ML/TAD), MIT (ML/ALS),
  - 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

# Échéancier

- 13h PA de Biologie
- 16h PA d'Informatique
- ... début des inscriptions ...
  - ↪ choix du PA, lettre de motivation, choix de cours
  - ↪ entretiens individuels
- 29 mai  $\approx$  fin des inscriptions et validation (si OK)
- réunion de rentrée  $\rightsquigarrow$  entretiens individuels (modif cours, projet, stage, 4A)
- amphis de présentation des stages BIO591 et INF591
- novembre : limite catégorie 1
- décembre : limite catégories 2 et 3, limite stages
- janvier : dernières candidatures étranger

à bientôt, j'espère