



# Master de Bioinformatique

Université de Rennes 1

[master-bioinfo@univ-rennes1.fr](mailto:master-bioinfo@univ-rennes1.fr)

**Descriptif des études - Maquette envisagée pour 2018-2019, sous réserve de validation par le conseil de l'UFR SVE et par la Commission de Formation de la Vie Universitaire**





## *FIXME : table des matières*

1. Le master en fonction de votre cursus
2. Le master 1
3. Le master 2
  - 3.1 Organisation
  - 3.2 Parcours ADG
  - 3.3 Parcours BIS
  - 3.4 Parcours IBI





## Le master bioinformatique en fonction de votre cursus

Le master de bioinformatique de l'Université de Rennes1 forme des **experts en sciences de données biologiques et génomiques** à l'interface de la santé, de l'environnement, des données massives et des technologies de l'information. La formation délivre un **socle de compétences communes en génomique et en informatique**. Dans ces deux domaines, la formation prend le temps d'introduire les notions utiles en fonction du cursus d'origine des étudiants. De nombreux stages et projets permettent d'approfondir ces notions. Nous encourageons les stages à l'étranger et la participation à des compétitions internationales.

Le master de bioinformatique est co-porté par les UFR SVE et Médecine. Il bénéficie d'un accord Erasmus+ avec l'université de Salerne (Italie), ainsi que d'une convention avec les filières Statistiques pour les sciences de la vie et Statistiques et ingénierie des données de l'ENSAI (les élèves ingénieurs de troisième année sélectionnés peuvent suivre respectivement les parcours BIS et IBI).

Le master de bioinformatique est ouvert aux étudiants biologistes, informaticiens, mathématiciens, bioinformaticiens (évidemment) ayant obtenu une licence ou un DUT, ainsi qu'aux étudiants en cursus santé, en parallèle avec leurs études de médecine. Nous accueillons également des étudiants au profil atypique ou en reprise d'étude via la formation continue.

### Parcours type Analyse de Données Génomiques

**Profil** : Analyser données \*omiques, utiliser infrastructures de calcul intensif

**Offres de stage** : très nombreuses (sur Rennes, en France et à l'étranger)

### Parcours type Bioinformatique en Santé

**Profil** (adapté aux cursus santé) : Aider au diagnostic, Génomique et cancérologie

**Offres de stage** : en augmentation

### Parcours type Informatique et Biologie Intégrative

**Profil** : Développer de nouveaux outils d'analyse pour les données biologiques

**Offres de stage** : nombreuses

M2 : Métagénomique @SVE	M2 : Machine learning @SVE, Médecine, Pharmacie	
M2 : Évolution moléculaire @SVE	M2 : Données massives et Web Sémantique @SVE, Médecine, Pharmacie	
M2 : Bio des syst. simul @SVE	M2 : Génétique somatique des cancers @Médecine	M2 : Bio des syst. simul @SVE
M2 : Molecular genetics		M2 : Bio des syst. réseaux @SVE
M2 : NGS en diagnostic @Médecine		M2 : Algorithmique des séquences @IRISA
M2 : Analyse de données NGS, cas pratiques @Médecine, Agrocampus		M2 : Méthodes d'optimisation @IRISA
M2 : Modélisation des macro-molécules (30h) @SVE		
M2 : Anglais (30h) @SVE, Médecine		
M1 : Bloc de compétences en génétique et génomique (180h) @SVE		
M1 : Bloc de compétences en informatique (200h) @SVE, Médecine		
M1 : Bloc de compétences transversales et stage (120h + stage) @SVE, Médecine		



## Vous êtes titulaire d'une licence de biologie, d'informatique, de bioinformatique, de statistiques ou de mathématiques

Vous pouvez candidater pour entrer en master 1

## Vous êtes étudiant(e) en cursus santé

Vous pouvez candidater pour effectuer votre master 1 en **parallèle à vos études de santé**.

- La version standard : les étudiant(e)s de médecine, pharmacie et odontologie valident leur master 1 à la fin de leur quatrième année après avoir validé les UE suivantes au cours de leur deuxième, troisième ou quatrième année :
  - Programmation impérative en Python (6 ECTS)
  - Bioinformatique et génomique 1&2 (6 ECTS)
  - Advanced statistics with R (3 ECTS)
  - Next Generation Sequencing (3 ECTS)
  - Bases de données relationnelles ( 6 ECTS)
  - Stage (6 ECTS)
- La version double cursus : les étudiant(e)s de médecine sélectionnés valident le master 1 en deux ans, pendant leur deuxième et leur troisième année. La première année est un tronc commun entre les masters portés par la faculté de médecine. La seconde année est spécifique à chaque master. Pour bioinformatique :
  - Bioinformatique et génomique 1 (3 ECTS)
  - Programmation impérative en Python (6 ECTS)
  - au choix
    - Advanced statistics with R (3 ECTS)
    - Next Generation Sequencing (3 ECTS)
  - Bases de données relationnelles ( 6 ECTS)
  - Stage (12 ECTS)

## Vous êtes élève-ingénieur(e) en seconde année à l'ENSAI

Vous pouvez candidater auprès de l'ENSAI pour bénéficier de l'OFPR au cours de votre troisième année afin d'obtenir à la fois votre diplôme d'ingénieur de l'ENSAI et le master de bioinformatique.

- Filière « Sciences de la vie » → parcours BIS (bioinformatique en santé)
  - ENSAI : 12 ECTS
    - Essais cliniques (6 ECTS)
    - Statistiques pour données \*omics (6 ECTS)



- Master bioinformatique : 18 ECTS
  - Bioinformatique et génomique 1&2 (6 ECTS)
  - Biologie des systèmes : réseaux biologiques (6 ECTS)
  - Gene mapping and NGS analysis (6 ECTS)
- Stage 6 mois (30 ECTS)
- Filière « Statistiques et ingénierie des données » → parcours IBI (info. et bio. Intégrative)
  - ENSAI : 12 ECTS
    - Big data (6 ECTS)
    - Statistiques pour la fouille et le big data (6 ECTS)
  - Master bioinformatique : 18 ECTS
    - Bioinformatique et génomique 1&2 (6 ECTS)
    - Biologie des systèmes : réseaux biologiques (6 ECTS)
    - Algorithmique des séquences (6 ECTS)
  - Stage 6 mois (30 ECTS)

### Vous êtes en reprise d'étude / en formation continue

Vous pouvez candidater en master 1 ou en master 2 en fonction de votre cursus et de l'expérience que vous avez acquise. Nous pouvons être amenés à vous proposer une sélection d'UE spécifique à votre situation, ainsi que d'étaler la durée de la formation sur plusieurs années. Merci de vous contacter directement ([master-bioinfo@univ-rennes1.fr](mailto:master-bioinfo@univ-rennes1.fr)).



## Le master 1 de bioinformatique





Le master 1 fournit un socle de compétences communes en génomique, NGS, génétique, programmation et gestion des données massives. Ces compétences permettent aux étudiants d'être autonomes dans la réalisation d'analyses bioinformatique ainsi que dans l'organisation et l'administration des solutions techniques.

Le master 1 est commun à tous les étudiants.

- Il commence par une session de mise à niveau début septembre, avec un approfondissement en génomique, informatique et/ou statistiques
- Au cours de l'année, les différentes UE concernent les trois groupes principaux de compétences :
  - **génétique et génomique** (180h)
  - **informatique** (200h)
  - compétences transversales et **insertion professionnelle** (120h)
- À partir d'avril, les étudiants peuvent s'engager dans un **stage (2 à 3 mois, prolongeable)**, participer à une **compétition internationale** ou s'engager dans un **projet personnel** (startup,...)

Bloc	UE	ECTS	Sem.
	Bootcamp/ Remise à niveau	0	S7
Stats	Apprentissages statistiques (Y. Le Cunff)	6	S7
Ins.Pro	Expression et Communication (A. Monnier)	3	S7
Géno	Bioinformatique en Génomique (C. Delamarche / A. Monnier)	6	S7
Géno	Analyse de Données Génomiques (A. Monnier)	6	S7
Prog	Systèmes Informatiques (E. Giudice)	3	S7
Prog	Programmation impérative Python (O. Dameron)	6	S7



Bloc	UE	ECTS	Sem.
 Anglais	English 1 (SCELVA)	3	S8
 Stats	Advanced Statistics with R (E. Becker)	3	S8
Ins.Pro	Stage ou projet ou compétition internationale (A. Monnier)	6	S8
 Géno	Next Generation Sequencing (A. Rolland)	3	S8
Géno	Genes and Genome Evolution (M. Ainouche)	6	S8
 Prog	Object Oriented Programming (E. Becker)	6	S8
Prog	Bases de données relationnelles (F. Moussouni)	3	S8







## Le master 2 de bioinformatique : parcours ADG, BIS et IBI

Le master 2 est ouvert à tous les étudiants ayant validé le master 1 de bioinformatique de Rennes. Nous accueillons également volontiers les candidatures extérieures d'étudiants d'autres masters de bioinformatique, de cursus santé, d'écoles d'ingénieurs etc. (après validation de leur dossier).

Vous devez choisir l'un des trois parcours :

- **Analyse de données génomiques (ADG)** se spécialise sur l'utilisation de méthodes de pointe pour l'analyse de données \*omiques, l'exploitation des bases de données de référence et l'utilisation d'infrastructures de calcul intensif. C'est la partie de la bioinformatique la plus proche de la biologie ;
- **Bioinformatique en santé (BIS)** se spécialise sur les enjeux actuels de la génétique et de la génomique en santé : aide au diagnostic, génomique et cancérologie, génétique de l'hérédité simple et complexe, ingénierie génétique, ainsi que sur la conception de thérapeutiques nouvelles et personnalisées. C'est la partie de la bioinformatique centrée sur les données biomédicales ;
- **Informatique et biologie intégrative (IBI)** se spécialise sur le développement de méthodes novatrices d'analyse de données biologiques grâce à des compétences en algorithmique et en programmation. C'est la partie de la bioinformatique la plus proche de l'informatique.

UE	ECTS	Parcours ADG	Parcours IBI	Parcours BIS
Modélisation des macro-molécules (E. Guidice)	3	x	x	x
English2 (SCELVA)	3	x	x	x
Gene mapping and NGS analysis (M. De Tayrac / S. Laguarrigue)	6	x		x
Evolution Moléculaire et Phylogénie (A. Salmon)	3	x		
Métagénomique (A. Dufresne / A. Quaiser)	3	x		
NGS en Diagnostic (C. Dubourg)	3	x		x
Molecular Genetics (L. Paillard)	6	x		x
Génétique Somatique des Cancers (M. De Tayrac)	3			x
Données Massives et Web Sémantique (O. Dameron)	3		x	x
Simulation des systèmes biologiques (S. Huet)	3	x	x	
Biologie des systèmes : réseaux biologiques (A. Siegel)	6		x	
Machine Learning for Biology (V. Monbet)	3		x	x
Algorithmique des séquences (P. Peterlongo)	6		x	
Techniques d'optimisation combinatoire (R. Andonov)	3		x	

### *M2 Parcours IBI : Informatique et Biologie Intégrative*

1. Biological Networks (6 ECTS, Anne Siegel)
2. Algorithmique des séquences (6 ECTS, Pierre Peterlongo)
3. Semantic Web and Big Data (3 ECTS, Olivier Dameron)
4. Machine Learning for Biology (3 ECTS, Valérie Monbet)
5. Techniques d'optimisation combinatoire (3 ECTS, Rumen Andonov)
6. Simulating Dynamic Systems (3 ECTS, Sébastien Huet)
7. Macromolecular Modeling (3 ECTS, Emmanuel Giudice)
8. Anglais (3 ECTS)

### *M2 Parcours ADG : Analyse de Données Génomiques*

1. Next Gen. Seq. : gene mapping and data analysis (6 ECTS, Marie de Tayrac et Sandrine Laguarrigue)
2. Next Gen. Seq. en diagnostic (3 ECTS, Christèle Dubourg)
3. Evolution Moléculaire et Phylogénie (3 ECTS, Armel Salmon et Malika Ainouche)
4. Analyses métagénomiques en écologie (3 ECTS, Alexis Dufresne et Achim Quaiser)
5. Ingénierie Génétique (6 ECTS, Luc Paillard)
6. Simulating Dynamic Systems (3 ECTS, Sébastien Huet)
7. Macromolecular Modeling (3 ECTS, Emmanuel Giudice)
8. Anglais (3 ECTS)

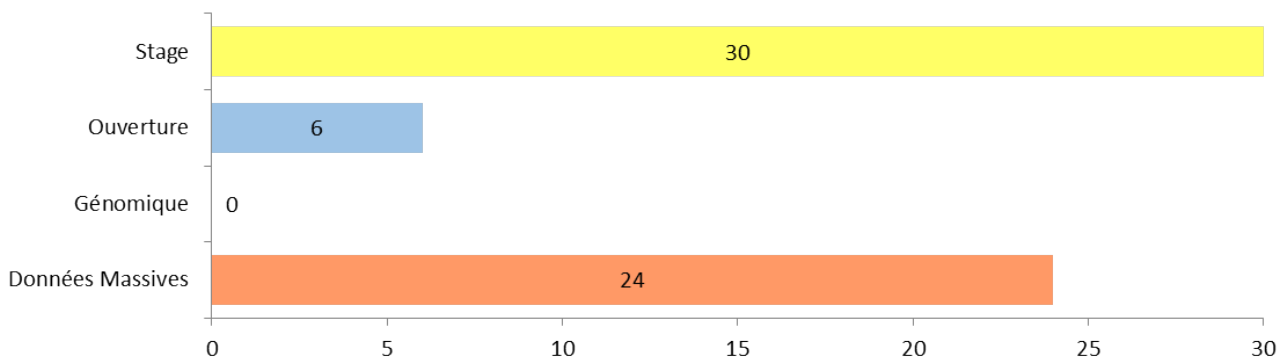
### *M2 Parcours BIS : BioInformatique en Santé*






1. Next Gen. Seq. : gene mapping and data analysis (6 ECTS, Marie de Tayrac et Sandrine Laguarrigue)
2. Next Gen. Seq. en diagnostic (3 ECTS, Christèle Dubourg)
3. Génétique Somatique des Cancers (3 ECTS, Marie de Tayrac)
4. Ingénierie Génétique (6 ECTS, Luc Paillard)
5. Semantic Web and Big Data (3 ECTS, Olivier Dameron)
6. Machine Learning for Biology (3 ECTS, Valérie Monbet)
7. Macromolecular Modeling (3 ECTS, Emmanuel Giudice)
8. Anglais (3 ECTS)



# Parcours IBI: Informatique et Biologie Intégrative

Mots clés: modélisation, données massives, machine learning, algorithmique



Bloc	UE	Crédits
 DM	Données Massives et Web Sémantique (O. Dameron)	3
 DM	Simulation des systèmes biologiques (S. Huet)	3
DM	Biologie des systèmes : réseaux biologiques (A. Siegel)	6
 DM	Machine Learning for Biology (V. Monbet)	3
Algo	Algorithmique des séquences (P. Peterlongo)	6
Algo	Techniques d'optimisation combinatoire (R. Andonov)	3
 Biophy.	Modélisation des macro-molécules (E. Guidice)	3
 Anglais	English 2 (SCELVA)	3





## Réseaux Biologiques

*Obligatoire pour les parcours : IBI*

*Volume horaire : 40h*

*Langue d'enseignement : Français*

*Responsable : Anne Siegel*

### Description

L'objectif du cours est d'acquérir différentes techniques pour analyser un réseau biologique à grande échelle, en particulier en le confrontant à différentes sources de données et d'observations hétérogènes sur l'organisme considéré. Les approches abordées couvriront l'étude des propriétés (statistiques) des graphes à grande-échelle et l'utilisation d'approches de vérification et de programmation par contraintes pour modéliser les informations portées par les jeux de données dans les langages expressifs.

Cours :

- Construction et analyse de réseaux à grande échelle, exemple d'application.
- Introduction à la programmation par contraintes : programmation par ensemble-réponse, optimisation linéaire.
- Intégrer des données hétérogènes et des connaissances.
- Utiliser de la programmation logique pour confronter des réseaux d'interactions à des données d'expression.
- Reconstruction et analyse de réseaux booléens.
- Reconstruction de cartes métaboliques pour intégrer des profilages métaboliques.

Travaux dirigés :

- 8 heures de TP, permettant de résoudre des questions biologiques présentées dans des articles scientifiques à l'aide d'outils d'intégration de données et d'analyses de réseaux. Les TP seront préparés et présentés par les étudiants eux-mêmes dans le cadre d'un travail préparatoire (mémoire).

## Algorithmique des séquences

Obligatoire pour les parcours : IBI

Volume horaire : 50h

Langue d'enseignement : Français

Responsable : Pierre Peterlongo

### Description

Cette UE propose une plongée dans l'algorithmique des séquences. L'objectif principal est double : fournir un sens critique aux utilisateurs d'outils bio-informatiques et offrir les clefs pour appréhender les algorithmes majeurs de ce domaine. Ainsi nous explorerons les algorithmes de *pattern matching*, d'alignement de séquence et nous aborderons les problématiques liées aux grandes masses de données NGS : structures d'indexation, *mapping* et assemblage de lectures. Ceci nous permettra également de comprendre et mettre en pratique les notions de complexité et de comprendre le concept d'heuristique et ses conséquences sur l'optimisation temps/mémoire et la qualité des résultats fournis.

Cours :

- Introduction à l'algorithmique du texte en suivant un fil rouge basé sur la notion de '*pattern matching*' dans les textes.
- Alignement de séquences et heuristiques associées, notions de précision et de sensibilité.
- Structures d'indexation de séquences (table de hachage, tableau et arbre des suffixes, transformée de Burrows Wheeler, graphe de De Bruijn).
- Algorithmes pour les données NGS : *mapping* et assemblage de génomes.

Travaux dirigés :

- Les TD suivent les cours magistraux. Ils consistent en des approfondissements des notions vues en cours par des mises en pratique : calcul de complexité, écriture d'algorithmes.
- Les TP ont lieu sur machine avec le langage Python. Ils servent à valider les approches, à les coder concrètement et à les tester sur des données réelles.

**Pré-requis : Algorithmique de base (boucles, récursivité, notions de complexité) et connaissance du langage Python. Connaissances de base en génomique et des données de séquençage NGS**



## Semantic Web and Big Data

*Obligatoire pour les parcours : IBI, BIS*

*Volume horaire : 20h*

*Langue d'enseignement : Anglais*

*Responsable : Olivier Dameron*

### Description

Cette UE porte sur les problèmes et les opportunités liées au partage, à l'intégration et à l'analyse de données massives en sciences de la vie en utilisant les techniques du Web Sémantique et de données liées (Linked Open Data).

L'enseignement aborde les données liées ((Open) Linked Data) ainsi que les technologies du W3C comme RDF et SPARQL qui permettent leur intégration et leur interrogation. Il présente ensuite le besoin d'interpréter ces données grâce à des connaissances symboliques formalisées par des ontologies. Là encore, ce volet s'appuie sur les technologies du W3C comme RDFS et OWL, ainsi que les capacités de raisonnement associées. L'enseignement porte également sur les principales ontologies biomédicales.

Compétences acquises :

- Annoter des données
- Représenter ces annotations en RDF
- Combiner ces annotations avec d'autres annotations sur des bases distantes en assurant l'interopérabilité
- Interroger des données RDF avec SPARQL
- Représenter des connaissances symboliques sous forme d'ontologies en RDFS et OWL
- Reasonner sur une ontologie en exploitant l'expressivité de son formalisme de représentation
- Reasonner sur des données en exploitant une ontologie
- Combiner plusieurs ontologies
- Connaître les principales ontologies biomédicales et les entrepôts comme BioPortal et [openlifedata.org](http://openlifedata.org)





## Machine Learning for Biology

Obligatoire pour les parcours : IBI, BIS

Volume horaire : 20h

Langue d'enseignement : Anglais

Responsable : Valérie Monbet

### Description

Cette UE transdisciplinaire originale regroupera des étudiants du master Bioinformatique, mais également des étudiants du master de Mathématiques, du Master de Calcul Scientifique (Physique), ou de Biologie Moléculaire et Cellulaire. L'objectif de l'UE est de mettre en œuvre les outils de Machine Learning, les valider, les comparer, et d'utiliser les modules Big Data/Apprentissage des logiciels R et/ou Python.

Cours :

- Modèle linéaire pour la régression
- Visualisation de données en grande dimension
- Méthodes linéaires pour la classification supervisée
- Comparaison, validation de modèles
- Agrégation de modèles, forêt aléatoires, boosting, ...
- Modèles graphiques
- Réseaux de neurones, deep learning
- Méthodes à noyau, SVM, ...

Pré-requis : Introduction aux statistiques (estimation), notions d'optimisation de fonctions convexes

## Techniques d'optimisation combinatoire appliquées à la bioinformatique

Obligatoire pour les parcours : IBI

Volume horaire : 18h

Langue d'enseignement : Français

Responsable : Rumen Andonov

### Description

L'objectif du cours est de se familiariser avec diverses techniques d'optimisation susceptibles d'être utilisées dans le domaine de la bioinformatique. Un accent particulier est réservé aux techniques de modélisation avec une prise en compte des conditions logiques. Une grande partie des méthodes présentées est basée sur la théorie de la Programmation Linéaire (PL) et les techniques de résolution des Programmes Linéaires en Nombres Entiers (PLNE). Ces méthodes, très développées aujourd'hui, permettent de concevoir des algorithmes efficaces pour la résolution de nombreux problèmes du domaine. La tâche est énormément facilitée aujourd'hui grâce à l'existence des solveurs contemporains comme CPLEX et GUROBI.

Cours :

- Techniques générales de modélisation, prise en compte de conditions logiques
- Programmation Linéaire (PL) et la méthode de simplex
- Techniques de base de la Programmation Linéaires en Nombres Entiers (PLNE)
- Applications :
  - Dans les réseaux métaboliques (le problème du flot compatible et ses variantes) ;
  - Dans le séquençage très haut débit (assemblage du génome).

Travaux dirigés :

- Les TD suivent les cours magistraux. Ils consistent en des approfondissements des notions vues en cours par des mises en pratique : calcul de complexité, écriture d'algorithmes.
- Les travaux pratiques permettront de modéliser certains des problèmes étudiés avec le langage AMPL et de les résoudre avec des solveurs contemporains comme CPLEX et GUROBI.

Pré-requis : Algorithmique élémentaire de niveau M1, notions de complexité algorithmique, notions de l'algèbre linéaire.

## *Simulating dynamic systems in Biology*

*Obligatoire pour les parcours : IBI, ADG*

*Volume horaire : 20h*

*Langue d'enseignement : Anglais*

*Responsable : Sébastien Huet*

### *Description*

La simulation *via* l'outil informatique de processus biologiques est maintenant présente dans la plupart des grands centres de recherche. Ce cours vise à donner une introduction aux principes de modélisation de systèmes dynamiques en biologie, centré sur les phénomènes à l'échelle cellulaire et tissulaire. Une part importante du cours sera dédié à la pratique, *via* le logiciel Python et l'utilisation de modules pré-existants. Les exercices d'applications seront directement issus de projet de recherche, en cours ou publiés.

Compétences acquises :

- Modélisation de phénomènes biologiques : comment passer de la description qualitative à la « mathématisation » ?
- Simulation informatique de ces phénomènes : comment reproduire l'expérience *in silico* ?
- Interprétation des modèles et des résultats des simulations
- Principe de l'analyse mathématiques des simulations
- Plus globalement, maîtrise minimale du langage Python et des modules dédiés à la simulation numérique

**Pré-requis : Intérêt pour les approches interdisciplinaires. Pas de pré-requis mathématiques spécifiques, toutes les notions nécessaires seront abordées ou rappelées au cours de l'UE.**



## *Macromolecular modeling*

*Obligatoire pour les parcours : IBI, ADG, BIS*

*Volume horaire : 20h*

*Langue d'enseignement : Anglais*

*Responsable : Emmanuel Giudice*

### *Description*

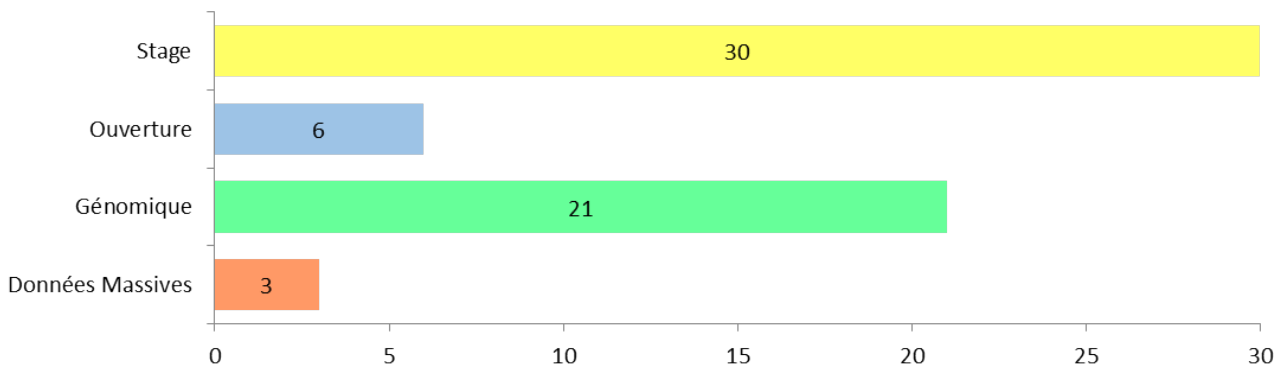
(nous sommes en attente du descriptif)










# Parcours ADG: Analyse de Données Génomiques

Mots clés: next generation sequencing, génomique, évolution, analyse de données



Bloc	UE	Crédits
 DM-BS	Simulation des systèmes biologiques (S. Huet)	3
 Géno	Gene mapping and NGS analysis (M. De Tayrac / S. Laguarrigue)	6
Géno	Evolution Moléculaire et Phylogénie (A. Salmon)	3
Géno	Métagénomique (A. Dufresne / A. Quaiser)	3
Géno	NGS en Diagnostic (C. Dubourg)	3
 Géno	Molecular Genetics (L. Paillard)	6
 Biophy.	Modélisation des macro-molécules (E. Guidice)	3
 Anglais	English 2 (SCELVA)	3



## Next Generation Sequencing : gene mapping and data analysis

Obligatoire pour les parcours : BIS, ADG

Volume horaire : 40h

Langue d'enseignement : Anglais

Responsable : M. de Tayrac, S. Laguarrigue

### Description

L'objectif de cette UE est de comprendre et maîtriser (manipulations) les concepts et outils liés au « gene mapping » en lien avec les données de génotypage et les notions de déséquilibre de liaison. Comprendre et maîtriser les outils de base d'analyse de données de séquençage. Annotation des variants : Etude de cas particuliers – observation, interprétation et mise en lien avec des phénotypes/maladies particulières. Comment aller plus loin dans la relation « génotype-phénotype » en intégrant des données hétérogènes de type Omiques

Cours :

- Partie Gene Mapping :
  - o Gestion de fichier de génotypage et contrôle qualité
  - o Description du déséquilibre de liaison
  - o Méthodes de Mapping
- Partie NGS and data Analysis :
  - o Déroulement de la procédure de traitement de donnée de séquences
  - o Réflexion sur les conséquences fonctionnelles des variants de type SNP/INDEL/ CNV
  - o Intégration de données hétérogènes de génomique
  - o Retour sur les concepts de génétique (editing, empreinte parentale, allèle spécifique expression...)

Travaux dirigés :

- Atelier NGS
- Analyse d'article.

## Next Generation Sequencing en diagnostic

Obligatoire pour les parcours : ADG, BIS

Volume horaire : 20h

Langue d'enseignement : Français

Responsable : Christèle Dubourg

### Description

L'objectif de cette UE est de comprendre les bases fondamentales de l'interprétation des mutations identifiées par séquençage nouvelle génération.

Cours :

- Expérience de panels de gènes en diagnostic à travers divers exemples (cancer, surcharge en fer...).
- Paysage mutationnel du syndrome de prédisposition aux cancers du sein et de l'ovaire à partir du séquençage haut débit d'un panel de gènes.
- Apport du séquençage massivement parallèle des gènes connus en pathologie dans le diagnostic de la déficience intellectuelle.
- L'exome en diagnostic.
- Dépistage non invasif des altérations chromosomiques.
- A quand le WGS en diagnostic ?
- Etude de l'impact des variants sur l'épissage par l'utilisation des minigènes.
- Tests fonctionnels basés sur le zebrafish.
- NGS en microbiologie.

Travaux dirigés :

- Atelier tri de variants issus d'un exome à partir de fichier VCF.
- Analyse d'article.





## *Evolution Moléculaire et Phylogénie*

*Obligatoire pour les parcours : ADG*

*Volume horaire : 20h*

*Langue d'enseignement : Français*

*Responsable : Armel Salmon*

### *Description*

- Introduction au raisonnement phylogénétique
- Modèles d'Evolution moléculaire
- Concepts et méthodes de reconstruction phylogénétiques :
- Arbres de distances (UPGMA, NJ)
- Méthode de Parcimonie
- Méthode de Maximum de Vraisemblance (ML)
- Analyse Bayésienne
- Conflits phylogénétiques
- Introduction à la Phylogénomique

Travaux dirigés :

- Exercices d'application sur les méthodes de reconstruction phylogénétiques à partir de différents jeux de données moléculaire



## *Analyses métagénomiques en écologie*

*Obligatoire pour les parcours : ADG*

*Optionnelle pour les parcours : -*

*Volume horaire : 30h*

*Langue d'enseignement : Anglais*

*Responsable : Alexis Dufresne*

### *Description*

(nous sommes en attente du descriptif)

## Ingénierie génétique

*Obligatoire pour les parcours : ADG, BIS*

*Volume horaire : 40h*

*Langue d'enseignement : Anglais*

*Responsable : Luc Paillard*

### *Description*

Our goal is to highlight how research projects in molecular genetics can be planned and organized. Several examples of research projects will be introduced as conferences. A focus will be made on how the projects were formalized as grant applications, taking into account the feasibility (cost, available resources...), the expected results, and the interest of the research for the funder.

"Witnesses" introduce their research thematics and provide essential information. Then, they explain what was they wrote in a grant application on this thematics, and how they conducted their research project. Which questions were raised, based on which hypotheses? Which methodologies did they propose, and why? What was their work plan? What did they do themselves, what was done in collaboration or outsourced, and why? What was done (or what is being done) after getting the grant, are there discrepancies between the initial project and the real life? Globally, witnesses provide any relevant information about how they conceived and managed a project.

Tutors similarly introduce their research thematics. Then, they focus on some pending questions, which might lead to a research project that can be discussed with the students.

### *Course goals*

Examples of topics include, but are not limited to, cohesinopathies, DNA replication, hormonal control of epigenetic modifications, epigenetic control of spermatogenesis, reticulum stress...

## Simulating dynamic systems in Biology

Obligatoire pour les parcours : IBI, ADG

Volume horaire : 20h

Langue d'enseignement : Anglais

Responsable : Sébastien Huet

### Description

La simulation *via* l'outil informatique de processus biologiques est maintenant présente dans la plupart des grands centres de recherche. Ce cours vise à donner une introduction aux principes de modélisation de systèmes dynamiques en biologie, centré sur les phénomènes à l'échelle cellulaire et tissulaire. Une part importante du cours sera dédié à la pratique, *via* le logiciel Python et l'utilisation de modules pré-existants. Les exercices d'applications seront directement issus de projet de recherche, en cours ou publiés.

Compétences acquises :

- Modélisation de phénomènes biologiques : comment passer de la description qualitative à la « mathématisation » ?
- Simulation informatique de ces phénomènes : comment reproduire l'expérience *in silico* ?
- Interprétation des modèles et des résultats des simulations
- Principe de l'analyse mathématiques des simulations
- Plus globalement, maîtrise minimale du langage Python et des modules dédiés à la simulation numérique

Pré-requis : Intérêt pour les approches interdisciplinaires. Pas de pré-requis mathématiques spécifiques, toutes les notions nécessaires seront abordées ou rappelées au cours de l'UE.



## *Macromolecular modeling*

*Obligatoire pour les parcours : IBI, ADG, BIS*

*Volume horaire : 20h*

*Langue d'enseignement : Anglais*

*Responsable : Emmanuel Giudice*

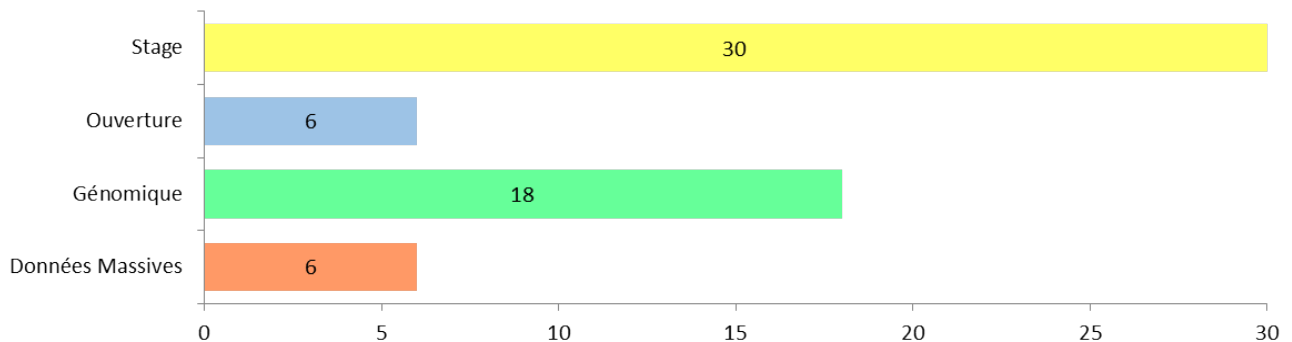
### *Description*







(nous sommes en attente du descriptif)



# Parcours BIS: Bioinformatique en Santé

*Mots clés: next generation sequencing, santé humaine, machine learning*



Bloc	UE	Crédits
 DM	Données Massives et Web Sémantique (O. Dameron)	3
 DM	Machine Learning for Biology (V. Monbet)	3
 Géno	Gene mapping and NGS analysis (M. De Tayrac / S. Laguarrigue)	6
Géno	Génétique Somatique des Cancers (M. De Tayrac)	3
Géno	NGS en diagnostic (C. Dubourg)	3
 Géno	Molecular Genetics (L. Paillard)	6
 Biophy	Modélisation des macro-molécules (E. Guidice)	3
 Anglais	English 2 (SCELVA)	3



## *Next Generation Sequencing : gene mapping and data analysis*

*Obligatoire pour les parcours : BIS, ADG*

*Volume horaire : 40h*

*Langue d'enseignement : Anglais*

*Responsable : M. de Tayrac, S. Laguarrigue*

### *Description*

L'objectif de cette UE est de comprendre et maîtriser (manipulations) les concepts et outils liés au « gene mapping » en lien avec les données de génotypage et les notions de déséquilibre de liaison. Comprendre et maîtriser les outils de base d'analyse de données de séquençage. Annotation des variants : Etude de cas particuliers – observation, interprétation et mise en lien avec des phénotypes/maladies particulières. Comment aller plus loin dans la relation « génotype-phénotype » en intégrant des données hétérogènes de type Omiques

Cours :

- Partie Gene Mapping :
  - Gestion de fichier de génotypage et contrôle qualité
  - Description du déséquilibre de liaison
  - Méthodes de Mapping
- Partie NGS and data Analysis :
  - Déroulement de la procédure de traitement de donnée de séquences
  - Réflexion sur les conséquences fonctionnelles des variants de type SNP/INDEL/ CNV
  - Intégration de données hétérogènes de génomique
  - Retour sur les concepts de génétique (editing, empreinte parentale, allèle spécifique expression...)

Travaux dirigés :

- Atelier NGS
- Analyse d'article.



## Next Generation Sequencing en diagnostic

Obligatoire pour les parcours : ADG, BIS

Volume horaire : 20h

Langue d'enseignement : Français

Responsable : Christèle Dubourg

### Description

L'objectif de cette UE est de comprendre les bases fondamentales de l'interprétation des mutations identifiées par séquençage nouvelle génération.

Cours :

- Expérience de panels de gènes en diagnostic à travers divers exemples (cancer, surcharge en fer...).
- Paysage mutationnel du syndrome de prédisposition aux cancers du sein et de l'ovaire à partir du séquençage haut débit d'un panel de gènes.
- Apport du séquençage massivement parallèle des gènes connus en pathologie dans le diagnostic de la déficience intellectuelle.
- L'exome en diagnostic.
- Dépistage non invasif des altérations chromosomiques.
- A quand le WGS en diagnostic ?
- Etude de l'impact des variants sur l'épissage par l'utilisation des minigènes.
- Tests fonctionnels basés sur le zebrafish.
- NGS en microbiologie.

Travaux dirigés :

- Atelier tri de variants issus d'un exome à partir de fichier VCF.
- Analyse d'article.



## Génétique somatique des Cancers

*Obligatoire pour les parcours : BIS*

*Volume horaire : 20h*

*Langue d'enseignement : Français*

*Responsable : Marie de Tayrac*

### *Description*

L'objectif de cette UE est de décrire et d'approfondir les grands champs de la génétique somatique comme outils de compréhension et de gestion de la pathologie tumorale. Cette UE vise particulièrement à permettre aux étudiants d'acquérir un regard critique sur les technologies et méthodologies innovantes utilisées aujourd'hui en génétique somatique et particulièrement autour des données de séquençage nouvelle génération.

Cours :

- Les particularités des génomes tumoraux
- Génomique et cancer
- Hétérogénéité clonale et perspectives évolutionnistes du cancer
- Génétique moléculaire des prédispositions au cancer
- Criblage mutationnel, apport dans la connaissance des tumeurs du rein
- Marqueurs compagnons et traitement des tumeurs solides
- Suivi des Maladies résiduelles en Onco-hématologie

Travaux dirigés :

- Atelier NGS et paysage mutationnel en cancérologie
- Analyse d'article.



## Ingénierie génétique

*Obligatoire pour les parcours : ADG, BIS*

*Volume horaire : 40h*

*Langue d'enseignement : Anglais*

*Responsable : Luc Paillard*

### *Description*

Our goal is to highlight how research projects in molecular genetics can be planned and organized. Several examples of research projects will be introduced as conferences. A focus will be made on how the projects were formalized as grant applications, taking into account the feasibility (cost, available resources...), the expected results, and the interest of the research for the funder.

"Witnesses" introduce their research thematics and provide essential information. Then, they explain what was they wrote in a grant application on this thematics, and how they conducted their research project. Which questions were raised, based on which hypotheses? Which methodologies did they propose, and why? What was their work plan? What did they do themselves, what was done in collaboration or outsourced, and why? What was done (or what is being done) after getting the grant, are there discrepancies between the initial project and the real life? Globally, witnesses provide any relevant information about how they conceived and managed a project.

Tutors similarly introduce their research thematics. Then, they focus on some pending questions, which might lead to a research project that can be discussed with the students.

### *Course goals*

Examples of topics include, but are not limited to, cohesinopathies, DNA replication, hormonal control of epigenetic modifications, epigenetic control of spermatogenesis, reticulum stress...

## *Semantic Web and Big Data*

*Obligatoire pour les parcours : IBI, BIS*

*Volume horaire : 20h*

*Langue d'enseignement : Anglais*

*Responsable : Olivier Dameron*

### *Description*

Cette UE porte sur les problèmes et les opportunités liées au partage, à l'intégration et à l'analyse de données massives en sciences de la vie en utilisant les techniques du Web Sémantique et de données liées (Linked Open Data).

L'enseignement aborde les données liées ((Open) Linked Data) ainsi que les technologies du W3C comme RDF et SPARQL qui permettent leur intégration et leur interrogation. Il présente ensuite le besoin d'interpréter ces données grâce à des connaissances symboliques formalisées par des ontologies. Là encore, ce volet s'appuie sur les technologies du W3C comme RDFS et OWL, ainsi que les capacités de raisonnement associées. L'enseignement porte également sur les principales ontologies biomédicales.

Compétences acquises :

- Annoter des données
- Représenter ces annotations en RDF
- Combiner ces annotations avec d'autres annotations sur des bases distantes en assurant l'interopérabilité
- Interroger des données RDF avec SPARQL
- Représenter des connaissances symboliques sous forme d'ontologies en RDFS et OWL
- Reasonner sur une ontologie en exploitant l'expressivité de son formalisme de représentation
- Reasonner sur des données en exploitant une ontologie
- Combiner plusieurs ontologies
- Connaître les principales ontologies biomédicales et les entrepôts comme BioPortal et [openlifedata.org](http://openlifedata.org)



## Machine Learning for Biology

Obligatoire pour les parcours : IBI, BIS

Volume horaire : 20h

Langue d'enseignement : Anglais

Responsable : Valérie Monbet

### Description

Cette UE transdisciplinaire originale regroupera des étudiants du master Bioinformatique, mais également des étudiants du master de Mathématiques, du Master de Calcul Scientifique (Physique), ou de Biologie Moléculaire et Cellulaire. L'objectif de l'UE est de mettre en œuvre les outils de Machine Learning, les valider, les comparer, et d'utiliser les modules Big Data/Apprentissage des logiciels R et/ou Python.

Cours :

- Modèle linéaire pour la régression
- Visualisation de données en grande dimension
- Méthodes linéaires pour la classification supervisée
- Comparaison, validation de modèles
- Agrégation de modèles, forêt aléatoires, boosting, ...
- Modèles graphiques
- Réseaux de neurones, deep learning
- Méthodes à noyau, SVM, ...

Pré-requis : Introduction aux statistiques (estimation), notions d'optimisation de fonctions convexes



## *Macromolecular modeling*

*Obligatoire pour les parcours : IBI, ADG, BIS*

*Volume horaire : 20h*

*Langue d'enseignement : Anglais*

*Responsable : Emmanuel Giudice*

### *Description*

(nous sommes en attente du descriptif)