**Bio-informatique**

Responsables du master : Annie Chateau, Séverine Berard. Enseignantes chercheuses en informatique.

**I ) Introduction**

**Les domaines liés** : Biologie, informatique, mathématiques, modélisation, statistique, physique, phylogénie, chimie, génomique, analyse d’image, médecine, biologie des populations, ADN, structure 3D, biochimie, biomathématique, protéomique, biophysique, ARN, cristallographie, séquençage, transcriptomique …

**Le contexte :** décrypter le vivant (schéma d’un chromosome)

**Les enjeux :**

Le traitement de l’information biologique :

* Stocker : acquisition et organisation (base de données)
* Analyser : modèles, méthodes, logiciels (donner du sens à l’information)
* Visualiser : gros volume de données

→ Pour produire de nouvelles connaissances

Les applications :

* Scientifiques :
  + Compréhension du vivant, modélisation d’écosystèmes
  + Fonctionnement des cellules
  + Evolution des espèces
* Economiques :
  + Médecine : maladie génétique, parasites, médicaments…
  + Biotechnologie, bioénergies, agro-alimentaire (OGM)
  + Environnement
* Sociétales :
  + Santé publique, nutrition
  + Propagation des maladies
  + Environnement

**II ) Quelques principes/notions**

**Le vivant :**

Une grande diversité, des principes universels :

* Cellule : unité constructive de base du vivant
* ADN : support de l’information génétique
* Protéines : macromolécules actives
* Dogme central de la biologie moléculaire : production des protéines à partie de l’ADN (via ARN)
* Code génétique : code de la traduction de l’ADN en protéine
* Le gène est l’unité de base de l’information génétique

**Le code génétique :**

* ADN : 4 lettres
* 20 protéines différentes
* On lit l’ADN par triplet (cordon)
* Combien de cordons différents ? 4^3 = 64
* Le code génétique est redondant

**L’évolution**

Basé sur l’histoire des organismes. 1ere forme de vie sur Terre : procaryote (bactérie) dans les milieux aquatiques en -4 md av JC. Ensuite, eucaryote. Ensuite Faune de Burgess, puis sortie de l’eau, puis mammifères, dinosaures, puis hominidés.

L’évolution est la transformation, des espèces vivantes au cours des générations. Ces changements successifs peuvent aboutir, à partir d’une seule espèce, à la formation de nouvelles « espèces-filles » = spéciations.

**La phylogénie**:

Conception d’un arbre de vie du XXe siècle par Ernst Haeckel.

Classification phylogénétique du vivant de 2006

**L’autre dogme centrale : la deduction par homologie**

L’évolution des gènes :

* Elle laisse des traces visibles lorsque l’on compare leur séquences.
* Évolution des gènes = substitution, insertions, délétions.
* Les gènes des organismes modernes sont issus de remaniements de gènes ancestraux : on peut déduire leur fonction par comparaison avec les gènes homologues d’autres espèces
* Il existe des gènes universels
* Homologie = parenté = ancêtre commun
* Pour une notion quantative, on parle de similitude, de ressemblance
* En pratique :
  + Une ressemblance importante est considérée comme une homologie
  + En dessous d’un certain seuil, il devient difficile de distinguer une homologie d'un ressemblance fortuite

**III ) Une question scientifique**

**Un exemple de questionnement** Les oiseaux sont

**La démarche : 3 problèmes**

* Comment obtenir les séquences d’ADN ? *Assemblage*
* Comment comparer ces séquences ? *Alignement*
* Comment reconstituer les relations de parenté ? *Phylogénie*

1. **Le problème de l’assemblage**

**Séquençage d’ADN**

* Récupération d’un échantillon de l’espèce
* Séquençage par une machine (amplification, casse aléatoire, traduction)
* En sortie, des millions de petits morceaux de génomes : les reads

Nécessiter d’assembler ces petits bouts

Heureusement, certains morceaux se chevauchent.

**Formalisation**

Donnée : Un ensemble F de séquences (les morceaux)

Résultats : Une séquence, la plus courte possible, contenant toutes les séquences de F (le génome)

* Problème informatique : « Plus courte super-chaine commune »
* Problème difficile (NP-complet)
* S’ajoutent d’autres difficultés : erreurs de lectures, existence de nombreuses régions répétées (entre 10 et 90%)

1. **Le problème de l’alignement**

L’alignement de séquences est la principale méthode de comparaison permet

* D’identifier des régions conservées, on en déduit l’homologie
* D’étudier les processus d’évolution à l’échelle moléculaire
* De déterminer la distance évolution et construire des phylogénies

Comparer des séquences serait relativement simple si elles avaient toutes la même longueur

Comme ce n’est pas le cas, il faut les aligner, c’est-à-dire trouver ou se situent les insertions et les délétions

**Qu’est ce qu’un alignement ?**

Mise en correspondance de 2 séquences.

3 opérations (les 3 événements évolutifs possible

**Ensemble des alignements**

Séquences X := a, Y := c à 3 alignements possibles

Séquences X := ac et Y := cg à 13 alignements possibles

1. **Le problème de la phylogénie**

Un arbre phylogénétique donne la configuration la plus probable pour rendre compte du degré de parenté existant entre les séquences.

Les

**VI ) Historique : quelques dates clé**

**La préhistoire**

* 1736 : Naissance de la théorie des graph

**Les début : 60-70 :**

* 1961 : Découerte du code génétique
* 1967 : « Construction of Phylogenetc Trees », Fitch & Margoliash
* 1968 : Atlas of Proteins

**L’accélération**

* 1985-88 : Programmes « Fasta » (Pearson-Lipman) & 1990 « Blast » (Altschul et al.) pour rechercher des séquences dans les banques
* 1990 : Ouverture du WorldWide Web
* 1994 : ClustalW, logiciel d’alignement multiple
* 1995 : 1er génome complet d’un organisme vivant Haemophilus influenzae
* 1996 : Génome de la levure
* 1990 à 2003 : Projet génome humain
* 2003 à aujourd’hui : projet ENCODE
* 2008 à aujourd’hui : mille génomes

**Aujourd’hui**

On the future of genomic Data

**Master Bio-informatique**

Master-bioinfo-contact@umontpellier.fr

**Master bio informatique de l’université de Montpellier – FDS**

Une formation avec des bases solides : La bio informatique au niveau Master depuis 2001 à Monptellier. C’est une formation reconnue, et un réseau d’anciens étudiants étendu en entreprises et en organisation de recherche.

Qui fait peau neuve : Passage en mention 2001 :

* Visibilité et stabilité
* Plus adaptée aux profils hétérogènes (Bio, Info, BioInfo)
* Ouverture en alternance (3 apprentis M1 et 1 apprentie M2 pour 2021)

Ouverture sur le monde professionnel :

* Rencontre université – entreprise
* Participation régulière à JOBIM
* Organisation des Montpellier Omics Days : rencontre de la communauté workshops ouverts à tous, implication des professionnels (Oxford Nanopore)

**Compétences**

* Modélisation informatique des problèmes biologiques / bio-informatiques (Spécification pluridisciplinaire, algorithmes classiques, notion de complexité)
* Conception et développement de solutions informatiques (Analyse UML, programmation objet/web/script, notions de système et réseau)
* Conception et exploitation

**Présentation du contenu**

Organisation de la formation :

* Origine des étudiants : Biologie, santé, informatique, è S1 adapté
* S2-S3 modules de bioinformatique : montée en puissance vers le cœur de métier
* UE professionnalisantes
* Nombreux stages et projets, Apprentissage par projet è compétences par la pratique

**Equipe pédagogique**

Une équipe soudée et variée, une trentaine d’enseignant(es) et d’intervenant(es) :

* Un noyau dur provenant de dépatrement informatique de la Faculté des Sciences (6 MCF en bioinformatique)
* 3 composantes
* 5 départements d’enseignement

**Données d’insertion des dipômées :**

**Secteurs professionnels concernés**

Tertiaire : Santé, Sciences, recherche et développement scientifique

Débouchés professionnels

**Emploi dans la région occitanie**

**Candidatures**

Conditions d’accès : être titulaire d’une licence d’informatique, de mathématiques, de biologie ou de bio-informatique, ou d’un diplôme équivalent.

Bassin de recrutement national + étude en France

>300 candidatures finalisées en M1 pour 30 places affichées en 2021 :

- Taux d’acceptation vers 25%

-

**Conclusion**

Une seule adresse mail à retenir : Master-bioinfo-contact@umontpellier.fr

Sites à visiter : Département informatique FDS, Offre de formation UM, Trouver mon master