

Journée des doctorants

Programme complet

9h30 **Introduction**

9h40 – 10h00 **Michaël BEKAERT** *Génétique Moléculaire de la Traduction, IGM*

Le recodage, *in silico* et *in vivo*

Les mécanismes non conventionnels de traduction (recodage) sont employés par certains virus comme le HIV ou le SRAS. En se basant sur l'organisation structurale des génomes, comment identifier de tels événements dans un génome eucaryote tel que celui de la levure *Saccharomyces cerevisiae*, et comment valider biologiquement ces prédictions informatiques?

10h10 – 10h30 **Jean-Paul FOREST** *Bioinformatique, LRI*

Modélisation des sites de décalage de cadre en -1

Le recodage lors de la traduction désigne l'ensemble des exceptions à lecture séquentielle du message porté par le brin d'ARN. Je présenterai tout d'abord le modèle consensuel des sites de décalage de cadre en -1, qui est une des formes de ce recodage. Je montrerai ensuite comment nous avons affiné ce modèle. Dans un troisième temps, j'expliquerai comment on peut détecter ces sites à partir de séquences primaires, que ce soit dans des génomes sauvages ou aléatoires. J'expliquerai également comment nous validons notre travail expérimentalement.

10h40 – 11h00 **Claire HERRBACH** *Bioinformatique, LRI*

Conception de matrices de substitution pour la comparaison de structures d'ARN

Comparer des molécules d'ARN nécessite de tenir compte de leurs différents niveaux de structures. Mon travail consiste à construire des matrices de substitution, par une étude statistique de séquences d'ARN et de leurs structures, à la manière de Henikoff et Henikoff pour les matrices BLOSUM. Jusqu'à maintenant, je n'ai pris en compte que les structures secondaires.

11h10 – 11h30 **Pause-café**

11h40 – 12h00 **David ABERGEL** *Bio-Informatique des Génomes, IGM et LIMSI*

Caractérisation des régions inter-ORF chez la levure *Saccharomyces cerevisiae*

Les régions inter-ORF sont impliquées dans le fonctionnement des gènes et la régulation générale du génome. Pour étudier les caractéristiques de ces zones à l'échelle génomique, nous avons utilisé et amélioré des méthodes bioinformatiques. Un outil se base sur les chaînes de Markov et l'autre outil utilise une modélisation de la trajectoire 3D de l'axe de l'ADN nu, à travers plusieurs modèles. Les résultats préliminaires seront discutés.

12h10 – 12h30 **Ana ARRIBAS GIL** *Probabilités, statistique et modélisation, LMO*

Mise en place d'un modèle évolutif pour l'analyse comparative des génomes

Dans le but de retrouver de petits motifs conservés en dehors des régions non-codantes de certains gènes, on s'est proposé d'identifier des plages de vitesse d'évolution homogène dans les séquences d'ADN (les vitesses d'évolution les plus lentes correspondant à des contraintes de sélection plus importantes et donc, à des séquences jouant un rôle important) à l'aide d'un modèle probabiliste d'évolution.

12h40– 14h00 **Déjeuner libre**

Journée des doctorants (suite)

14h00 – 14h20 **Joan HERISSON** équipe Geste et Image, LIMSI

Visualisation et Exploration 3D de l'ADN à grande échelle : un nouveau point de vue pour l'analyse des génomes

Les séquences d'ADN sont le plus généralement représentées sous forme textuelle : une succession de lettres (A, C, G et T). Des algorithmes très efficaces ont été développés sur ces séquences textuelles depuis plusieurs années. Cependant, ces séquences sont des signaux mono-dimensionnels alors que l'ADN est une structure tri-dimensionnelle ; une partie de l'information contenue par les molécules d'ADN n'est donc pas représentée par ces séquences. La visualisation 3D permet d'une part de représenter les 3 dimensions de l'ADN et offre d'autre part un point de vue global sur celui-ci. Ainsi, grâce au logiciel ADN-Viewer développé au LIMSI, un utilisateur peut observer et explorer un chromosome entier d'ADN nu.

14h30 – 14h50 **Quentin SCULO** *Evolution Moléculaire et Génomique, IGM*

Méthodes de construction d'arbres phylogénomiques de procaryotes

La comparaison exhaustive des protéines codées par les génomes procaryotes entièrement séquencés permet d'estimer une distance évolutive séparant chaque organisme. Plusieurs méthodes ont été mises au point pour calculer ces distances et en déduire une matrice permettant de construire un arbre retraçant la phylogénie de 80 organismes qui représentent les grands embranchements des procaryotes.

15h00 – 15h20 **Julie BERNAUER** *Génomique Structurale, LEBS*

Modélisation des complexes protéine-protéine

Dans le cadre du projet de génomique structurale de la levure, nous nous intéressons aux complexes protéine-protéine potentiels et à leur modélisation. Nous avons pour objectif la réalisation d'un programme de *docking*. La partie géométrique de l'algorithme utilise les diagrammes de Voronoï des partenaires qui permettent de simplifier l'exploration de l'espace. La construction de ces diagrammes sur les complexes dont la structure est connue expérimentalement permet, par ailleurs, de mettre au point une fonction de score statistique.

15h30 – 15h50 **Yann PONTY** *Bioinformatique, LRI*

Modélisation de structures génomiques, génération aléatoire et applications

A partir d'outils classiques de l'analyse combinatoire, on étudie les différents formalismes permettant de décrire les structures de séquences d'intérêt en génomique. Ces études permettent alors l'élaboration d'algorithmes de générations aléatoires, dont les séquences produites pourront être utilisés pour la validation d'un modèle.

16h00 – 16h20 *Pause-café*

Journée des doctorants (suite et fin)

16h30 – 16h50 **Sarah COHEN BOULAKIA** *Bioinformatique, LRI*

Intégration d'Informations pour l'étude des tumeurs cancéreuses : le projet HKIS

Le projet européen HKIS (5ème PCRD, IST) vise à la constitution d'une plate-forme intégrée de stockage et de traitement de données biologiques et biomédicales en cancérologie. Les enjeux d'un tel projet sont multiples et se placent notamment au niveau de la gestion de l'hétérogénéité des données. Après avoir présenté HKIS, nous dégagerons plus précisément les problèmes auxquels nous avons à faire face pour intégrer les données nécessaires à l'étude des cancers et nous présenterons le cadre suivi dans le projet pour les uniformiser et permettre leur interopérabilité.

17h30 – 17h20 **Mathieu ROCHE** *Inférence et Apprentissage, LRI*

EXIT : EXtraction Itérative de la Terminologie

Les travaux que nous présentons consistent à extraire la terminologie à partir de textes spécialisés et notamment à partir d'un corpus de résumés d'articles relatifs à la Biologie Moléculaire. Notre approche, en partie fondée sur des méthodes statistiques, consiste à extraire, de façon itérative, des termes binaires (par exemple "single strand"). Ceci permet de construire de manière itérative des termes de longueur plus importante particulièrement pertinents dans le domaine de la biologie (par exemple "single strand DNA binding protein"). Quelques exemples de termes sont consultables sur la page Web : <http://www.lri.fr/ia/Genomics>.

17h30 – 17h50 **Romain RIVIERE** *Bioinformatique, LRI*

Titre non communiqué

Résumé non communiqué.

18h00 ***Discussion générale et conclusion***

-
- Institut de Biotechnologie des Plantes ([IBP](#))
 - Institut de Génétique et Microbiologie ([IGM](#))
 - Laboratoire d'Enzymologie et Biochimie Structurales ([LEBS](#))
 - Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur ([LIMSI](#))
 - Laboratoire de Mathématiques d'Orsay ([LMO](#))
 - Laboratoire de Recherche en Informatique ([LRI](#))