



Le réseau régional d'ingénieurs en bioinformatique de Lille et le PPF bioinformatique vous convient **lundi 14 Mai 2012** à une conférence sur l'utilisation des outils GMOD (Generic Model Organism Database).

Elle se déroulera de 13h30 à 17h30 dans l'amphithéâtre de l'Institut de Biologie de Lille  
1, rue du Pr Calmette, LILLE.

#### Première partie

**13.30 – 15.00 : Olivier Arnaiz - CGM**

**Centre de Génétique Moléculaire UPR 3404 Gif-sur-Yvettes**

Le projet GMOD : une collection d'outils génériques pour le stockage et la visualisation de données biologiques



**15.00 – 15.30 : Pause**

#### Deuxième partie

**15.30 – 17.00 : Joelle Amselem**

**INRA - URGI / BIOGER Versailles**

Genome annotation : URGI pipelines and manual annotation system



Contact : [sophie.gallina@univ-lille1.fr](mailto:sophie.gallina@univ-lille1.fr)



**Olivier Arnaiz CGM Centre de Génétique Moléculaire UPR 3404 Gif-sur-Yvettes**

## **Le projet GMOD : une collection d'outils génériques pour le stockage et la visualisation de données biologiques**

L'objectif Les techniques modernes, en biologie (séquençage, transcriptome, ...), génèrent aujourd'hui des quantités de données de plus en plus importantes. Et la plupart des petites communautés n'ont pas les moyens matériels/humains de recréer leur propre système de gestion de données. De plus le développement d'une nouvelle base de données communautaire représente un coût non négligeable. De cette idée est né le projet GMOD pour « Generic Model Organism Database ».

L'objectif du projet GMOD est donc de fournir une série d'outils génériques, clé en main, pour gérer et visualiser différents types de données biologiques. Derrière le projet GMOD, des centaines de développeurs repartis dans le monde améliorent les logiciels et aident les utilisateurs à leurs mise en place. Aujourd'hui plus d'une centaine de bases de données utilisent les outils GMOD.

Chado et GBrowse sont deux des logiciels les plus populaires du projet GMOD. Chado est un schéma modulaire de base de données relationnelle. Il est capable d'intégrer la plupart des types de données biologiques rencontrés aujourd'hui (génomique, transcriptomique, génétique,...). Le GBrowse est une interface web interactive permettant la visualisation d'annotations sur un génome. Grâce à sa très grande flexibilité et sa configuration relativement simple, GBrowse est probablement le navigateur de génome le plus utilisé.

Je vous présenterai ces deux outils dans le cadre de ParameciumDB, une base de données biologiques que nous développons.

## Joelle Amselem INRA - URGI / BIOGER Versailles

### Genome annotation : URGI pipelines and manual annotation system

J. Amselem<sup>1,2</sup>, N. Lapalu<sup>1,2</sup>, B. Brault<sup>1,2</sup>, L. Brigitte<sup>1,2</sup>, A. Keliet<sup>1</sup>, E. Kimmel<sup>1</sup>, D. Steinbach<sup>1</sup>, M.H. Lebrun, H. Quesneville<sup>1</sup>

<sup>1</sup>INRA, URGI, Versailles, France

<sup>2</sup>INRA, BIOGER, Thiverval Grignon, France

contact : [joelle.amselem@versailles.inra.fr](mailto:joelle.amselem@versailles.inra.fr)

With the recent development of new generation sequencing technologies, a large number of genomes are sequenced, producing very large amount of data. This increasing amount of data has to be stored, analyzed and searched. To face this challenge, the URGI platform (<http://urgi.versailles.inra.fr>) provides tools to store and annotate entirely sequenced genome such as annotation pipelines, databases and user-friendly interfaces to browse and query the data.

I will focus here on the distributed annotation system. This system relies on the well known GMOD databases (BioSeqFeature and chado schema) and interfaces (Gbrowse, Gbrowse\_Syn and Apollo). The manual gene curation system allows to create/curate and validate gene structure. Data are available and shared by the consortium community as soon as they are committed in the database using the “pure JDBC” direct communication protocol between Apollo and Chado. Quick and advanced search are proposed, respectively through GnpIS QuickSearch (<http://urgi.versailles.inra.fr/gnpis>) and BioMart (complex search criteria on Datamart, GMOD Tool). BioMart results could be exported to a Galaxy server for further bioinformatic analysis. We will present these tools in the frame of fungal genome sequencing projects

I also will present here Transposable element detection and annotation pipelines TEde novo and TEannot from REPET package [1].

We will present this integrated genome annotation system successfully used for fungal genomes such as *Botrytis cinerea* T4 [2] and *Leptosphaeria maculans* [3] in the frame of their genome sequencing project. Portals for the different species are available at <http://urgi.versailles.inra.fr/index.php/Species>.

#### References:

- [1] T. Flutre, E. Duprat, C. Feuillet and H. Quesneville, Considering transposable element diversification in de novo annotation approaches. PLoS ONE 6(1):e16526, 2011
- [2] Amselem J, Cuomo CA, van Kan JA, Viaud M, et al, Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. 2011 Aug;7(8):e1002230. Epub 2011 Aug 18.
- [3] Rouxel T, Grandaubert J, Hane JK, Hoede C, et al. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat Commun. 2011 Feb 15;2:202. doi: 10.1038/ncomms1189.