

Compte rendu réunion 1 GUGGO

Groupe de travail des utilisateurs Galaxy du Grand Ouest GUGGO : Galaxy User Group Grand Ouest

Réunion : 5 avril 2012 – IRISA

Présents

Site de Rennes (par ordre du tour de table) Olivier Collin, Cyril Monjeaud, Claudia Hériveau, Anthony Bretaudeau, Jérôme Montfort, Fabrice Legeai, Dominique Lavenier, Aurélien Roult, Olivier Sallou, Stéphanie Mottier, Marc Aubry, Thomas Derrien, Nicolas Soriano, Christophe Caron, Yann Audic, Pierre Peterlongo

Site de Nantes (par ordre du tour de table) Audrey Bihouée, Dominique Tessier, Raluca Teusan, Julien Gras, Edouard Hirchaud, Eric Charpentier

Excusé

Gilles Lassalle

Inventaire des actions « Galaxy » en cours dans le Grand Ouest

GenOuest

Une instance expérimentale a été mise en place à l'adresse galaxy.genouest.org. Cette instance accepte les connexions en

provenance d'utilisateurs disposant déjà d'un compte sur les machines de la plate-forme (connexion via LDAP).

Etant donné le caractère expérimental, très peu d'outils ont été installés.

D'autre part, une machine virtuelle basée sur BioLinux et possédant une instance Galaxy a été créée. Cette image peut être utilisée sur un poste de travail pour réaliser des premiers tests et notamment mettre en place des workflows.

La procédure d'acquisition d'une machine Galaxy avec un espace disque conséquent est en cours (obtention récente des devis Dell).

Institut de Génétique et Développement de Rennes

S. Mottier, Yann Audic

Pour pouvoir traiter des données de RNAseq fournies par le Genoscope, une instance Galaxy a été installée sur un poste local avec divers outils de base. Des problèmes de configuration ont été rencontrés avec les outils Tophat et Cufflinks

Plate-forme de génomique des insectes

F. Legeai

Une instance locale a été mise en place pour des partenaires INRA en France. Des outils de base tels que Bowtie et FastQC ont été installés. Cette instance est en phase de test avec quelques utilisateurs. Il est prévu pour assurer une montée en charge de migrer ensuite vers la plate-forme GenOuest.

ABiMS: Station Biologique de Roscoff

L'environnement Galaxy fait l'objet d'expérimentations depuis

2010. Un serveur opérationnel est en place depuis quelques semaines sur une machine dédiée (64 Go mémoire) et capable de soumettre des travaux sur le gestionnaire de tâches SGE.

Les travaux de développement de pipelines seront focalisés sur des thématiques de la plate-forme de métabolomique Corsaire. L'environnement Galaxy sera également mis à profit pour accompagner les projets Investissement d'Avenir hébergés par la Station Biologique de Roscoff.

Pour les utilisateurs extérieurs à la Station Biologique il est possible de demander de compte pour accéder à la ressource Galaxy. La procédure est disponible à l'adresse : <http://abims.sb-roscoff.fr/>

BiRD Nantes

Pas d'expérience mais la plate-forme est intéressée par l'outil pour mise à disposition de certaines ressources NGS.

Retour d'expérience sur l'utilisation de Galaxy

Institut de Génétique et Développement de Renne

Des instances de Penn State et Tübingen ont été utilisées pour des analyses RNASeq. La question qui se pose, plus que l'environnement de traitement, est de savoir comment efficacement traiter les données. Un échange d'expérience au sein d'une communauté serait un plus.

BIA Nantes

Utilisation d'une instance publique pour l'analyse de données NGS. Cette instance, bien que limitée à 3Go a permis d'expérimenter. Une stratégie originale consiste à payer à un

accès à un cloud avec les outils complets afin de bénéficier de l'intégralité de la richesse de l'offre puis de passer ensuite en local.

INRA Rennes

Durant la discussion on évoque H. Falentin, utilisatrice des outils de Galaxy

Synthèse générale sur l'utilisation

L'outil est très apprécié car il offre un environnement prêt à l'emploi. Les tutoriaux fournis sur le site central sont très bien faits. Il est possible d'envoyer des résultats d'analyse sur des Brower (UCSC) et ainsi visualiser. Cette fonctionnalité est très appréciée par les scientifiques travaillant sur des génomes disponibles et séquencés. La situation est différente pour les autres communautés.

La richesse de l'environnement est à la fois sa richesse et sa faiblesse. Il faut en effet une première connaissance des outils. Il y a beaucoup de capsules pas forcément faciles à appréhender.

Pour certaines tâches plus complexes, il n'y a pas l'outil dans Galaxy et il faut repasser dans l'environnement Linux. Il est par contre possible de rajouter des outils. Certains participants évoquent des difficultés pour rajouter des outils (descripteur XML et programme Python).

Rubrique people

Anton Nekrutenko, un des chefs de projet de Galaxy, va faire un séjour à Lyon en tant que chercheur invité. <http://lbbe.univ-lyon1.fr/-Nekrutenko-Anton-.html> On peut envisager la possibilité de l'inviter à venir donner une conférence ou un séminaire sur Galaxy.

Formations

Les formations proposées sur l'environnement pourraient être de deux types : – journée générale sur l'utilisation de l'outil : maîtrise de l'environnement et de sa philosophie – ensuite des journées d'approfondissement sur certains outils.

Plan d'action

La mise en place d'une instance pour la communauté du Grand Ouest. Installation des outils essentiels et des workflows adaptés.

On pourrait par exemple, envisager des instances dédiées : métagénomique, NGS, etc. Remarque : Cette solution n'est peut-être pas la plus efficace car elle pourrait obliger l'utilisateur à passer d'une instance à l'autre pour trouver certains outils. Il serait probablement plus efficace d'arriver à une classification claire des différents outils.

Les catégories envisageables :

- génomique : polymorphisme
- expression : épissage
- métagénomique : ADN, ARN

Un plan d'action en plusieurs étapes est envisageable pour la mise en place d'une ressource commune :

- Mise en place d'un portail Galaxy sur GenOuest et ABiMS
- Mise en place de formations sur l'environnement. Ces formations permettent de rassembler une communauté d'utilisateurs
- Sondage auprès de la communauté d'utilisateurs préalablement constituée afin d'établir la liste des logiciels prioritaires

Parallèlement à cette mise en place d'autres actions sont mises en place pour enrichir l'offre :

- Réflexion sur les avantages de l'outil ToolShed pour une communauté Biogenouest avec le partage d'outils entre ABiMS/BiRD et GenOuest. q
- Interfaçage d'outils NGS de l'équipe GenScale

Prochaine réunion du groupe

A l'issue de l'installation de l'instance GenOuest et de la mise en production de l'instance ABiMS pour coordonner les actions de formation.