Compte Rendu – 15/02/2016

Rencontre plateau bioinformatique – équipes

# Bilan

L'ensemble des équipes ont été rencontrées, pour certaines très au complet, d'autres un peu moins. Il se peut donc qu'il y ait des besoins ou des possibilités d'interactions pas encore envisagées. Chaque équipe a présenté ses projets en détaillant les thématiques, les données, les protocoles, les moyens humains et parfois la valorisation associée ou envisagée.

Toutes les équipes n'ont pas forcément un besoin fort de support. Pour ces équipes, un fonctionnement en mode échange et discussions sera certainement suffisant lorsque les analyses sont maîtrisées en interne. Pour d’autres équipes le support attendu est plus sous forme d’un accompagnement à l’analyse avec donc la nécessité d’interagir sur les protocoles et de former sur les outils. Pour finir, certaines équipes sont en attente d’un support fort, notamment dans la réalisation d’analyses. Dans ce dernier cas, un fonctionnement type partenaire de projet doit se mettre en place pour définir la réalisation des tâches.

Pour résumé, on distingue 3 profils :

* Equipe indépendante (maîtrise des protocoles, utilisation serveurs calculs)
* Equipe avec moyen humain, mais aide et support à mettre en place. Interactions sur les protocoles, ainsi que formation spécifiques. Eventuel besoin d’utilitaires à développer en ligne de commande ou clic-bouton (galaxy)
* Equipe avec peu de moyen humain, investissement à prévoir pour la réalisation des analyses ou encadrement fort de stagiaire, personnel non-permanent.

# Thématiques

Ci-dessous une tentative de tableau récapitulatif des grandes thématiques abordées par les équipes.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | AMAR | AMPC | Botrytis | Epidemio | EGIP | Lepto | RSYST |
| Polymorphisme (SNP + Indels) | X |  | X | X | X | X |  |
| Variants Structuraux |  |  |  |  | X |  |  |
| Duplications segmentales |  | X |  |  |  |  |  |
| QTL |  |  |  | X |  | X |  |
| GWAS |  |  |  | X | X |  |  |
| RNA-Seq | X | X | X |  | X | X | X |
| RNA-Seq (mi-RNA) |  |  | X |  |  |  |  |
| Chip-Seq |  | X |  |  |  | X |  |
| MAINE-Seq |  |  | X |  |  | X |  |
| Assemblage (génome) |  |  | X |  | X | X |  |
| Assemblage (transcriptome) |  |  |  |  |  |  |  |
| Annotation structurale |  |  |  |  | X |  |  |
| Annotation fonctionnelle |  | X | X |  | X |  |  |
| G. comparative (orthologues) |  | X |  |  |  | R.Dean ? |  |
| Protéomique / Spectro de masse | X |  |  |  |  | X |  |
| Identification (amplicon) |  |  |  |  |  |  | X |
| Métagénomique |  |  |  |  |  | X | X |
| Métatranscriptomique |  |  |  |  |  | X | X |

# - Propositions générales sur le mode de fonctionnement

Ci-dessous une proposition de mode de fonctionnement, ainsi que les différents investissements de l’équipe sur les thématiques des équipes. Cette proposition correspond à l’état des lieux fait au 15/02/2016. En fonction de l’évolution des thématiques, ainsi que des projets, cela sera actualisé pour prioriser de nouveaux besoins.

## Communication

Nous souhaitons attirer l’attention sur la communication à destination de la cellule bioinformatique concernant les différentes analyses bioinfo prévues dans les projets. Impliquer la cellule au début des projets peut nous permettre d’être force de proposition sur les moyens, protocoles ou encore interlocuteurs préférentiels (via réseau bioinfo) pour la réalisation du travail.

## Fonctionnement / Ressources

Le suivi des tâches et travaux réalisés par le plateau bioinfo sera tracé dans le gestionnaire de tâche Redmine. Afin de pouvoir faciliter les interactions ainsi que le suivi du travail, il sera demandé aux équipes de l’utiliser afin de consigner le travail sur le projet correspondant. Le but est de garder trace du travail réalisé par chacun.

Nicolas a 100% de son temps affecté au plateau, il s’investira donc potentiellement sur tous les aspects traités. Adeline a environ 20% de son temps affecté au plateau, elle garde donc une grande partie de son temps pour l’analyse des données Botrytis. Son investissement côté plateau portera donc préférentiellement sur les tâches ou thématiques transversales ou encore les outils spécifiques du plateau (portail web, …).

La cellule bioinfo possède une machine pouvant accueillir des programmes spécifiques et être ouverte sur demande pour des tests ou des jeux de données de taille raisonnable. Pas de stockage définitif. N’hésitez pas à demander un accès si nécessaire.

Nous encourageons préférentiellement l’utilisation de serveurs institutionnels (Genotoul, URGI, Geneouest…), pour lesquels nous proposons un accompagnement à l’utilisation.

Pour le stockage pérenne des données brutes nouvellement acquises (ou des anciennes qui pourraient être ré-utilisées), nous préconisons l’utilisation de l’instance ARCHIVE du CATI BBRIC.

Il ne faut pas hésiter à nous solliciter sur des questions d’analyse ou plus orienté bio-informatique. Le bureau est ouvert en permanence et toute interaction est la bienvenue !

## Encadrements / Nouveaux arrivants

Il nous est tout à fait possible de participer à l’encadrement de stagiaire et autre, avec un investissement plus ou moins important selon la thématique. Cela s’accompagne aussi de la formation ou démonstration des outils de la cellule (gestionnaire de projet, ..) pour les nouveaux arrivants dans l’unité.

## Formation / Transfert compétences

Nous proposons de réaliser des formations sur demande selon les besoins des différentes équipes ou personnes. Afin d’être le plus efficace possible, les formations envisagées seront dans un premier temps très pratiques. Beaucoup de formations existent sur la bioinfo et l’analyse de NGS (alignement, assemblage, …), les langages de programmation (Python, R, …). Il faut que le besoin soit clairement identifié pour que l’on puisse définir si nous pouvons apporter un complément cohérent à une formation déjà existante. En revanche, nous réaliserons au fur et à mesure des supports de formations et tutoriels qui seront mis à disposition (ex : utilisation ordonnanceur cluster calcul, transfert données fastq serveur archive CATI). Le but est de constituer un ensemble de ressources disponibles pour l’unité. Par la suite cela pourra s’agrémenter de protocole d’analyse ou autre utilitaire (ressource externe formation bioinfo, CATI).

Actuellement, les formations que nous pouvons organiser (uniquement sur demande) :

* Formation à l’utilisation basique de UNIX/Linux
* Formation à l’utilisation d’un cluster de calcul (ordonnanceur de job).
* Formations du CATI BBRIC (module à rejouer)
* Formation au versioning de code, sécurisation et partage de code informatique. ( à destination des personnes développant un peu de code, R,…).
* Formation à la virtualisation simple pour la mise en place d’environnement d’analyse. (à destination des personnes qui réalisent des analyses sur leur machine, comment isoler l’environnement de travail).

# - Propositions d’actions à court terme (6 mois +)

* Formation à l’utilisation de Redmine : réalisation d’une table ronde ITA formation sur l’outil, la prise en main, utilisation basique, …
* Mise en place d’une veille active et réalisation de protocoles d’analyse autour du RNA-Seq type « Dual RNA-Seq », biblio, tests : le but est de former un petit groupe de personnes travaillant sur des données RNA-Seq de type interaction hôte-pathogène et/ou cinétique pour définir des protocoles d’analyses (plan exp, données attendues, analyse diff, enrichissement). Cette thématique est très fortement ressortie et le besoin d’analyse est clairement identifié.
* Support à l’équipe AMPC, réalisation d’analyse de données sur le nouveau génome pacBio colleto (duplication segmental, génomique comparative, visualisation)
* Accompagnement Ralph Dean lors de son séjour : analyse de génomique comparative à définir. Quel support ? Quels besoins ?
* Suite suivi du projet Gandalf (Lepto, Epidemio)
* Outil d’extraction de séquence (gène dans différentes souches). Besoin exprimé par plusieurs équipes, cas d’utilisation à définir.

# - Propositions d’actions à long terme (1 an +)

La cellule bioinfo propose la mise en place d’un portail web pour mettre à disposition de l’ensemble de l’unité toutes les informations nécessaires.

* Retrouver les documents des présentations et formations de la cellule
* Des modes d’emploi (par exemple soumission données brutes à ARCHIVE)
* Des dossiers proposant des solutions pour des types d’analyse données, avec les liens vers les outils préconisés
* Des liens vers des sites utiles + les modalités d’accès (outils, formations, plateformes…) (ex : CATI, genotoul, divers GALAXY…)
* Accès aux outils développés par la cellule
* Liens vers les BD référentes pour chaque champignon de BIOGER + liens vers des bases plus généralistes
* Centraliser les localisations des bouquins bioinfo dans l’Unité

Réflexion à conduire autour de la mise à disposition des données, pérennisation, outil de visualisation (genome browser). Il n’y a pas de proposition claire sur le stockage des données finalisées. Il apparaît intéressant de conduire une réflexion sur quelles données ? où ? pour qui ?... Formation d’un petit groupe de personnes, définition claire des besoins, recherches de solution existantes, mise en place. Réunion à prévoir dans quelques mois.

# - Perspectives / Réflexion

Mettre en place des outils de gestion des données phénotypiques ? Mise en relation simplifiée des données omics / phénotypiques ?

Assemblage de transcrits ? isoformes : exploiter les données RNA-Seq disponibles pour tenter la recherche de variants alternatifs/ isoformes.