



## 4<sup>ème</sup> Assemblée Générale « orientée biologistes » Formation: métagénomique, épigénétique et visualisation 19-20 avril 2016 – Paris (UPMC) (réunion préparatoire le 18 avril à Paris)

**Membres du CATI présents le 18 (16):** Corinne Rancurel, Martine Da Rocha (Sophia), Martial Briand (Angers), Bernhard Gschloessl, Franck Dorkeld (Montpellier), Fabrice Legeai, Anthony Bretaudeau, Olivier Filangi (Rennes), Joseph Tran (Saclay) Adeline Simon, Nicolas Lapalu (Versailles), Patrice Baa-Puyoulet (Lyon) Sébastien Carrere, Ludovic Legrand, Ludovic Cottret, Jérôme Gouzy (Toulouse).

**Participants à la formation des 19/20 (10):** PONTS Nadia (Bordeaux), NEGRE Nicolas (Montpellier), JACQUIN-JOLY Emmanuelle (Paris), DARRASSE Armelle (Angers), LEBRETON Lionel, LIEGARD Benjamin, JUBAULT Melanie (Rennes), FUDAL Isabelle, LAVAL Valérie (Versailles), WARNEYS Romain (Angers)

*Les noms des personnes qui ont suivi une précédente réunion/formation « orientée biologistes » (2013-14-15) sont soulignés*

**Communication/inscriptions :** Arnaud Ridet (dpt SPE)

**Annonce :** <http://cati-bbric.toulouse.inra.fr/lib/exe/fetch.php/bbric-ag-communaute-201604-19-20.pdf>

**Liste de diffusion :** [communaute-bbric@liste.inra.fr](mailto:communaute-bbric@liste.inra.fr)

**Rédacteur :** Jérôme Gouzy

**Type du document :** compte rendu, relevé de conclusion et perspectives. Les documents de la formation sont disponibles sur [https://listes.inra.fr/sympa/d\\_read/communaute-bbric/Formation-2016/](https://listes.inra.fr/sympa/d_read/communaute-bbric/Formation-2016/)

## Planning des Assemblées Générales BBRIC

- 30 octobre 2012 : présentation des membres du CATI
- 15-16 avril 2013 : « orientée biologistes », présentation du CATI à notre « communauté servie »
- 28-30 octobre 2013 : « orientée technique informatique », présentation d'outils et de méthodes pour le développement. Travaux pratiques sur le déploiement de programmes via Galaxy.
- 23-25 avril 2014 : « orientée biologistes », présentation de l'architecture bioinformatique BBRIC, formation à l'utilisation de l'environnement BBRIC. Thèmes : analyse de données génomes et transcriptomes
- 23-24 octobre 2014 : « orientée technique (bio)informatique », présentation d'outils et de méthodes (bio)informatiques.
  - 20-22 octobre : Couplée à l'AG, Sébastien a organisé avec l'aide de la FP de Toulouse, une formation aux technologies web : HTML5, CSS3, JAVASCRIPT.
- 27-29 avril 2015 : « orientée biologistes », formation des utilisateurs de la communauté servie à l'utilisation de l'environnement BBRIC. Thèmes : annotation fonctionnelle & détection polymorphisme
- 28-30 octobre 2015 : « orientée technique (bio)informatique », présentation d'outils et de méthodes (bio)informatiques.
  - 26-27 octobre : Couplée à l'AG, LudoC a organisé avec l'aide de la FP de Toulouse, une formation de 3j à l'utilisation de la librairie javascript D3.js
- 19-20 avril 2016 : « orientée biologistes », formation des utilisateurs de la communauté servie à l'utilisation de l'environnement BBRIC. Thèmes : métagénomique, épigénétique et visualisation, présentation de la technologie PacBio
- Automne 2016 : « orientée technique info/bioinfo », présentation d'outils et de méthodes (bio)informatiques.

## Objectifs des formations BBRIC

1. Illustrer à travers des exemples variés comment nous allons interagir avec les biologistes
  - à distance et sur un temps long
  - à travers quelques principes de fonctionnement
  - grâce aux outils que nous mettons à disposition (Portail bioinformatique <https://bbric.toulouse.inra.fr>, Archive, Workspace, Galaxy, Serveur Web ad hoc, SMRT Portal pour l'analyse de données PacBio, etc.)
2. Rendre autonomes les utilisateurs sur les tâches d'analyse de données récurrentes et automatisées.
3. Illustrer un savoir faire bioinformatique pour la conception de pipeline d'analyses bioinformatiques

4. Produire un support de formation que nous pourrions réutiliser facilement
5. Essayer d'être complémentaire des autres formations, déjà disponibles ou en préparation.

## Retours sur la Formation 2015

Au total 45 personnes ont suivi des formations que nous avons organisées

- 16 personnes ont suivi la formation initiale à Paris en avril 2015
- 16 octobre à Toulouse (LIPM, LudoL, LudoC, Sébastien, Erika), 13 personnes (panachage modules 2014 et 2015)
- 12-13 octobre à Sophia (ISA, Martine & Corinne), 9 personnes
- 15-16 octobre à Sophia (ISA, Martine & Corinne), 7 personnes (modules de 2014)

C'est moins qu'en 2014 où 91 personnes avaient pu assister à la formation parisienne et aux répliques sur les centres. La principale raison de cette baisse est causée par l'investissement important nécessaire pour reproduire localement les modules.

## Planning de la formation 2016

Lundi 18 : Répétition de la partie théorique de la formation entre membres du CATI

### Mardi 19

Module	Horaire	Durée
Introduction et présentation du portail BBRIC	10H00-11H30	1 h 30
<b>Visualisation de résultats d'analyses</b>		
Créer son propre génome browser	11H30-12H45	1h15
<b>Repas - déjeuner</b>		<b>1h30</b>
Visualisation des réseaux	14h15-15h15	30mn
Visualisation d'informations hiérarchisées	15h15h-15h45	30mn
<b>Metagénomique</b>		
Caractérisation de métagénomomes à partir d'amplicons 16S et gyrB	16h-17h45	1h45
<b>Technologie PacBio</b>		
La technologie PacBio pour l'assemblage de génomes bactériens et eucaryotes. Présentation de SMRT Portal (1 <sup>ère</sup> partie)	17h45-18h30	45mn

## Mercredi 20

Module	Horaire	Durée
<b>Metagenomique</b>		
Assemblage de métagénomés bactériens	09h00-11h00	2h
<b>Pause</b>		
<b>Epigénétique</b>		
Analyse du méthylome bactérien avec des données PacBio et l'environnement d'analyse « SMRT portal »	11h15-12h15	1h
<b>Repas – déjeuner</b>		
Identification des régions régulatrices par une approche FAIRE-seq	13h45-15h15	1h30
<b>Discussion: bilan de la formation et identification des futurs besoins</b>	15h15-16h	45mn

## Organisation de la formation

La formation a été organisée par Martine Da Rocha (Sophia) qui a suivi la formation de formateur interne et nous guide dans l'application de la méthode recommandée. Pour rappel la méthode est séquencée en trois parties « Découverte/Démonstration/Application », elle nous donne un cadre commun qui favorise à la fois l'homogénéité de notre support pédagogique et la reproductibilité de la formation. Chaque module fait intervenir au moins 3 membres du CATI : l'expert qui a développé le pipeline, le responsable qui prépare le support pédagogique et présente le module, et le relecteur qui vérifie le support pédagogique et depuis cette année, prépare aussi le TP.

Le résultat 2016 est un support de plus de 260 diapositives incluant les objectifs pédagogiques de chaque module, la partie théorique et les questions pour les travaux pratiques. Ce support très complet sera réutilisé par ceux qui vont reproduire tout ou partie de la formation sur leurs sites respectifs (déjà prévu à Toulouse et à Sophia)

Pour tenir compte des commentaires de l'an passé, nous avons diminué le nombre de modules pour pouvoir donner plus de temps à la partie « Application/TP ». En outre, afin de mieux répartir l'effort au sein du trinôme qui prépare le module, cette partie TP a été préparée par le « Relecteur » et non plus par le responsable du module.

Comme les années passées, la journée de préparation en conditions réelles la veille de la formation nous permet d'affiner notre discours et de partager nos outils et méthodes entre bioinformaticiens du CATI. Comme notre AG technique de l'automne, c'est un moment indispensable pour favoriser le développement de compétences partagées au sein de notre réseau.

Grâce au soutien financier du SPE, nous avons loué à nouveau la salle de l'UMPC (1500€) pendant les vacances de la région parisienne. La prestation est vraiment très professionnelle et cela participe à la réussite de la formation. Comme les années passées, l'organisation pédagogique et pratique (avec les « formalités » qui vont avec) a nécessité un travail important de la part de Martine et Ludovic Legrand s'est à nouveau beaucoup investi pour que tous les pipelines soient fonctionnels à temps. Bravo et merci à tous les membres du CATI qui ont conçu et mis en œuvre cette formation pour nos collègues biologistes. Le tableau ci-après récapitule le rôle de chacun.

Responsable: Martine Da Rocha (ISA/Sophia)

Assistants: Ludovic Legrand (LIPM/Toulouse), Jérôme Gouzy (LIPM/Toulouse) & Arnaud Ridet (SPE/Sophia)

Modules	Responsable et intervenant principal	Expert	Relecteur
Introduction et présentation du portail BBRIC	Jérôme Gouzy, Sébastien Carrere, Ludovic Legrand (LIPM) Martine Da Rocha (ISA) Anthony Bretaudeau (IGEPP)		
Créer son propre génome browser	Anthony Bretaudeau (IGEPP)	Sébastien Carrere (LIPM)	Martial Briand (IRHS)
Visualisation des réseaux	Ludovic Cottret (LIPM)	Ludovic Cottret (LIPM)	Patrice Baa-Puyoulet (BF2I)
Visualisation d'informations hiérarchisées	Sébastien Carrere (LIPM)	Corinne Rancurel (ISA)	Franck Dorkeld (CBGP)
Caractérisation de métagénomés à partir d'amplicons 16S et gyrB	Bernhard Gschloessl (CBGP)	Martial Briand, Matthieu Barret (IRHS)	Olivier Filangi (IGEPP)
La technologie PacBio pour l'assemblage de génomes bactériens et eucaryotes. Présentation de SMRT Portal	Jérôme Gouzy (LIPM)	Jérôme Gouzy (LIPM)	Ludovic Legrand (LIPM)
Assemblage de métagénomés bactériens	Fabrice Legeai (IGEPP)	Bernhard Gschloessl (CBGP)	Nicolas Lapalu (BIOGER)
Analyse du méthylome bactérien avec des données PacBio et l'environnement d'analyse « SMRT portal »	Adeline Simon (BIOGER)	Ludovic Legrand (LIPM)	Corinne Rancurel (ISA)
Identification des régions régulatrices par une approche FAIRE-seq	Martine Da Rocha (ISA)	Fabrice Legeai (IGEPP)	Joseph Tran (IPS2)

## Feedback des personnes formées

Afin de recueillir les premiers retours des personnes ayant assisté à la formation, nous nous appuyons sur la discussion de fin de formation ainsi que sur le petit questionnaire rempli par nos collègues biologistes. Les réponses au questionnaire se trouvent à l'url [http://cati-bbric.toulouse.inra.fr/lib/exe/fetch.php/bbric-ag-communaute-201604-19-20-resultats\\_questionnaire\\_evaluation\\_biologistes.pdf](http://cati-bbric.toulouse.inra.fr/lib/exe/fetch.php/bbric-ag-communaute-201604-19-20-resultats_questionnaire_evaluation_biologistes.pdf)

Cette année encore les commentaires sont positifs. La seule petite réserve concerne le temps consacré aux TP. Nous y avons consacré beaucoup plus de temps mais pour certains modules ce n'est toujours pas suffisant.

Six des dix personnes avaient déjà suivi nos formations. Cela montre que ceux qui font l'effort de venir ont l'impression d'acquérir des connaissances utiles en génomique et dans l'utilisation d'outils bioinformatiques. Lors de notre discussion finale, ils ont confirmé l'intérêt de faire des formations sur plusieurs thèmes plutôt qu'une formation sur un seul thème. Peut être que cette organisation est plus adaptée aux collègues curieux par rapport à ceux qui souhaitent régler leur problème immédiat dans l'analyse d'un type de données particulier ?

Le point le plus positif concerne les nouveautés pédagogiques de l'année. Ainsi le nouveau module « myGenomeBrowser » que l'on joue en premier, permet d'une part de voir comment créer son génome browser avec ses données sans dépendre d'un bioinformaticien. Et d'autre part, il nous permet d'introduire une présentation des formats les plus usuels en bioinformatique, utilisés tout le long de la formation. Ainsi, les fiches « formats » que nous distribuons durant ce module ont été un support très utile pour tous les TP. Nos modules orientés « visualisation » où nous présentons des outils qui sont ensuite réutilisés lors des modules ultérieurs ont aussi été très bien accueillis. Pour finir, les fiches de synthèses utilisées comme support de discussion afin de valider l'acquisition des connaissances à la fin de chaque module ont été un outil pédagogique très efficace.

## Autoanalyse sur l'action de formation du CATI BBRIC

Cet outil d'interaction avec la « communauté servie » par le CATI était un projet structurant du CATI. A l'issue des 4 ans initialement prévus pour le CATI il est temps de faire le bilan sur cet aspect.

Durant 4 ans, nous avons interagit avec la communauté servie par le CATI. Par rapport au nombre important de sollicitations tout au long de l'année, peu de chercheurs se sont donné la peine de venir à nos assemblées générales et aux formations organisées à Paris. Néanmoins, la communauté BBRIC est bien plus large que les seuls biologistes assistant aux réunions/formations sur Paris. Nous interagissons au quotidien avec les biologistes de nos laboratoires et lors de nos réunions/formations nous avons pu interagir avec certain(e)s plus ou moins isolés qui avaient des idées, une stratégie et de la curiosité. Nous avons essayé de les aider en planifiant nos développements et nos formations pour les accompagner au mieux.

On note que le public qui participe aux répliques des formations sur les centres est un peu différent de celui qui assiste à la formation sur Paris. Ainsi, sur nos centres ce sont aussi bien des chercheurs que des techniciens qui assistent aux sessions alors que ce sont principalement des chercheurs qui participent aux formations sur Paris.

Tous les ans nous avons appliqué de nouvelles idées d'améliorations organisationnelles et pédagogiques et elles ont toutes amené une avancée. Il y aura toujours des moyens de s'améliorer mais grâce à l'organisation originale que nous avons mise en place, nous avons pu acquérir des compétences, une méthode et une expérience pédagogique que nous n'avions pas forcément au début du CATI.

Nous avons désormais un capital d'environ 850 diapositives support de formation sur les thèmes les plus importants de la génomique développé dans nos laboratoires. Nous devons améliorer en continu le support, en parallèle de nos outils, mais nous avons déjà un capital qui nous permettra de couvrir une bonne partie des demandes de notre communauté (cf ci-dessous la fiche récapitulant les « BBRIC protocols »).

Nous avons appris à construire ensemble des modules de formation, nous avons appris à partager nos travaux d'un point de vue technique et méthodologique et donc à travailler ensemble.

Notre organisation est lourde, l'effort fourni par les participants a été important mais le bilan aussi bien par rapport à l'objectif de former nos collègues que pour développer notre collectif informatique, objectif tout aussi important que le premier, est positif.

## Perspectives

Comme les années passées, nous avons profité de la rencontre avec des biologistes qui ne sont pas de nos laboratoires pour affiner notre perception des thèmes sur lesquels il y a une attente en termes de formation. Ainsi, les thèmes d'épigénétique, de visualisation et de métagénomique évoqués par les biologistes les années précédentes ont été abordé cette année. D'autres thèmes déjà mentionnés l'an dernier : phylogénie et génomique des populations sont revenus. Et comme nouvelle demande est apparue l'intégration de données hétérogènes (hors visualisation) et plus particulièrement les données RNA/Chip/FAIRE/\*/-seq.

En termes de pédagogie, certains pensent que se tourner vers la construction de MOOC permettrait que notre effort soit moins important. Le gain en termes d'effort et d'efficacité pédagogique n'apparaît pas à tous mais comme il y aurait une salle à Montpellier qui permettrait de faire des MOOC, à charge de Bernhard, Franck et Fabrice de faire un « MiniMooc » sur un module de leur choix afin de nous faire partager l'expérience lors de la prochaine AG d'automne.

On verra bien ce que donnent les tests de « virtualisation » mais lors des discussions nous sommes quand même assez d'accord sur le fait que l'organisation actuelle nous permet de garder un contact privilégié avec les apprenants et que c'est une occasion de rencontres et de partage de connaissances avec les autres membres du CATI.

Un compromis pourrait être de se filmer pendant la session de formation. Ainsi avec l'aide d'un film, il serait beaucoup plus facile pour d'autres de reproduire la formation sur les centres car une difficulté de la réplique reste de s'approprier les modules sur lesquels on n'est pas intervenu.

Nous avons aussi évoqué la possibilité de mixer formation sous forme de modules comme actuellement et des présentations sur des thèmes plus nouveaux ou exploratoires (comme nous l'avons un peu fait avec PacBio cette année).

Quoi qu'il en soit, nous n'avons pas de visibilité sur le devenir des CATI au delà de la fin de l'année, nous rediscuterons des prévisions lors de notre AG d'automne ... si l'on a des informations d'ici là.



## « BBRIC protocols » - version du 4 mai 2016

### Protocoles mis à disposition de la communauté BBRIC pour l'analyse de données

Le CATI BBRIC a développé et met à disposition de la communauté (« servie » par le CATI) un certain nombre de protocoles d'analyse de données génomiques. Le présent document inventorie les problèmes pour lesquels nous vous proposons un outil pouvant vous aider dans votre analyse. Nos protocoles sont disponibles via une interface web (basée ou pas sur Galaxy). Les protocoles pour lesquels nous avons dispensé une formation et pour lesquels nous disposons d'un support de formation sont également identifiés. Les supports sont disponibles dans les documents partagés de la liste de diffusion <https://listes.inra.fr/sympa/info/communaute-bbric> (envoyez un message à [communaute-bbric-request@listes.inra.fr](mailto:communaute-bbric-request@listes.inra.fr) pour vous inscrire)

#### **Problématiques abordées** (*attention, nous n'avons pas la prétention de traiter tous les cas d'utilisation*)

- **Assemblage**
  - Assemblage de petits génomes (<100Mb)
    - Données Illumina - **F2014**
    - Données PacBio – **F2016**
  - Suppression de la redondance dans un assemblage de transcrits
- **Annotation structurale de génomes**
  - Bactériens - **F2014**
  - De champignons
  - D'oomycètes
- **Annotation fonctionnelle de protéomes**
  - Basée sur le contenu en domaines fonctionnels – **F2015**
  - Par transfert à partir d'une espèce modèle – **F2015**
  - Identification de peptides signaux potentiels
  - Identification de candidats effecteurs (oomycètes)
- **Analyse du polymorphisme**
  - Détection de SNPs à partir d'un génome de référence - **F2015**
  - Détection de SNPs sans génome de référence - **F2015**
  - Prédiction de l'effet d'un polymorphisme - **F2015**
  - Construction d'une matrice de fréquences alléliques - **F2015**
- **Metagénomique**
  - Amplicons 16S et gyrB - **F2016**
  - Assemblages de métagénomes bactériens – **F2016**
- **Génomique comparative**
  - Détection de transferts horizontaux chez les métazoaires - **F2014**
  - Analyse de familles de protéines– **F2015**
- **Transcriptomique**
  - Mesure de l'expression sur la base d'un génome annoté - **F2014**
  - Assemblage de transcriptome avec mesure de l'expression - **F2014**
- **Métabolisme**
  - Identification de voies métaboliques - **F2015**
- **Visualisation**
  - De réseau - **F2016**
  - D'informations hiérarchisées - **F2016**
- **Couteau suisse**
  - Conversion de formats et soumission de génomes annotés aux banques publiques – **F2014**
  - Construction d'un serveur « blast » avec vos données
  - Construction d'un génome browser avec vos données – **F2016**
- **Protocole pour le contrôle qualité**
  - Dans le cadre d'une manip de RNAseq - **F2015**
  - Lors de l'assemblage de transcriptomes – **F2015**
- **Epigénétique**
  - Méthylome bactérien / PacBio - **F2016**
  - Identification de régions régulatrices en FAIREseq – **F2016**

Point d'entrée : <https://bbric.toulouse.inra.fr/> (login/password = login/password LDAP)