

La plateforme bioinformatique Migale

Assemblée générale du PEPI IBIS



Missions

- Mettre à disposition une **infrastructure de calcul scientifique** pour la génomique
- Diffuser un **savoir-faire en bioinformatique et biostatistique**
- Concevoir et développer des **applications**
- **Analyser** des données génomiques

Missions

- Mettre à disposition **une infrastructure de calcul scientifique** pour la génomique

- Infrastructure hébergée au DataCenter INRA Île de France :
 - 300 To de stockage
 - 1081 cœurs
 - 150 outils
 - 85 banques de données
- ~500 utilisateurs actifs

Missions

- **Diffuser un savoir-faire en bioinformatique et biostatistique**

- Cycle de formation à la bioinformatique par la pratique
 - 15 modules
 - 24 jours cumulés
 - 110 personnes formées

Missions

➤ Concevoir et développer des applications

- Mini-projets & projets de développement
 - Bases de données
 - Interface d'exploration de données génomiques
 - Développements & mise disposition de composants Galaxy (métagénomique, protéomique, ...)
- ~6 projets / an

Missions

➤ Analyser des données génomiques

- Collaboration / Accompagnement en
 - Génomique bactérienne
 - Génomique comparée
 - Métagénomique gène marqueur(s)
 - Métagénomique shotgun
 - Métatranscriptomique
- ~25 projets / an

Caractéristiques de la plateforme

- Ouverte à l'ensemble de la communauté des sciences du vivant
- **Infrastructure Scientifique Collective INRA**
- Plateforme certifiée **ISO9001** depuis 2011
- Membre de **France-Génomique** et de l' **Institut Français de Bioinformatique**

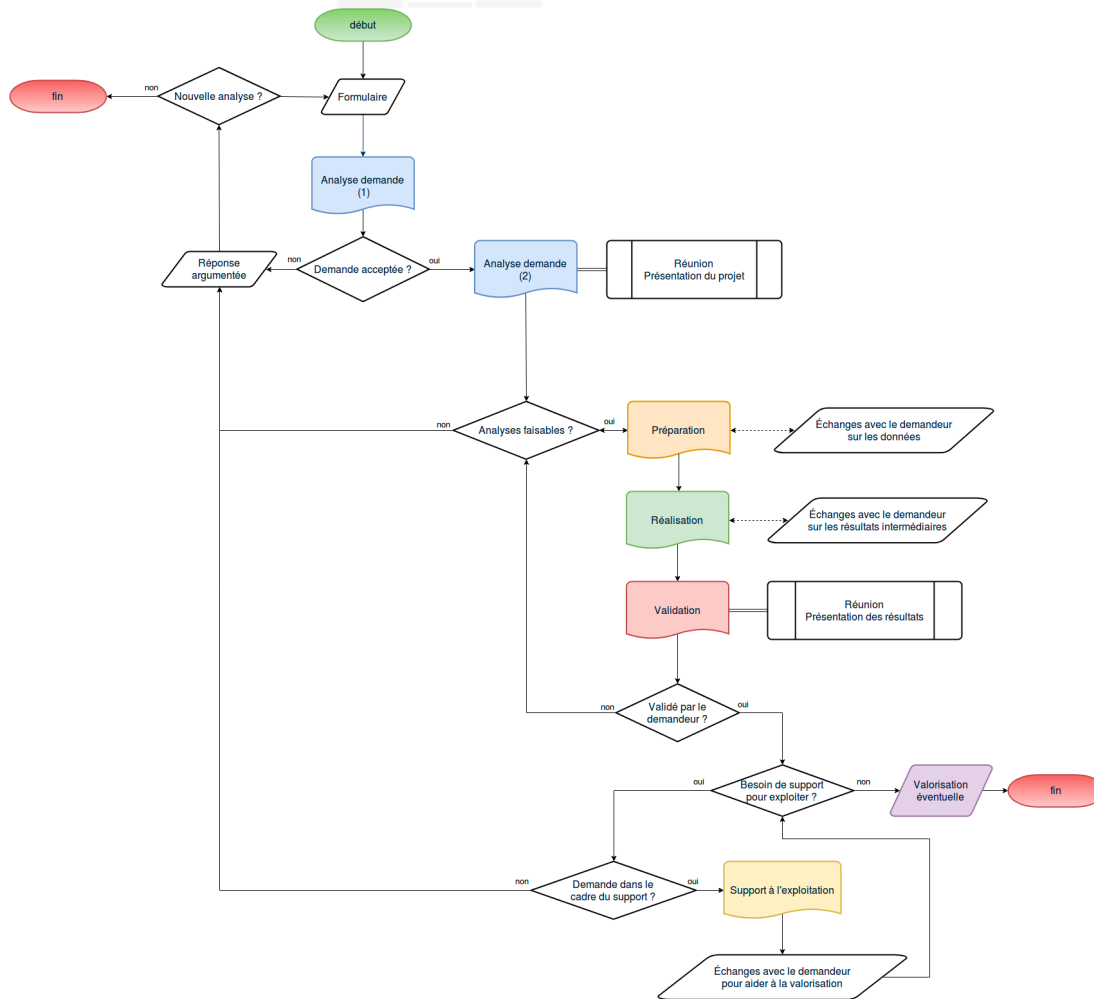


Focus : Analyse de données génomiques

- **Domaine d'expertise :**
 - **Génomique bactérienne** (QC, assemblage, annotation de génomes)
 - **Génomique comparée**
 - **Métagénomique** (barcoding / shotgun)
 - **Métatranscriptomique**
- **Périmètres du services :**
 - **Collaboration**
 - On s'occupe de traiter vos données
 - **Accompagnement :**
 - On se tient à votre disposition pour :
 - Conseils pour le plan d'expérience, le séquençage, les analyses à effectuer, les outils à privilégier...
 - Vous débbuger et vous aider à interpréter les résultats que vous obtenez

<http://genome.jouy.inra.fr/analyses/PROJECTS/report.html#>

Processus d'analyse de données génomiques



Étapes :

- **Expression des besoins**
- **Faisabilité**
- **Engagements (réciproques) sur les délais**
- **Phase de réalisation**
- **Validation par le demandeur**
- **Phase support**

Exemple : projet MetaPDO-Cheese



Partenaires

- INRA (GMPA, URF, MICALIS, MaIAGE)
- CNAOL
- France-Génomique, Genoscope

Objectifs

Montrer la **diversité microbienne** entretenue par la famille des **44 fromages AOP** français.
Comprendre les tenants de la **structuration des écosystèmes microbiens fromagers**.

Mise en œuvre

Analyse globale des flores des **laits et fromages** par **métabarcoding** (bactéries et levures)
Analyse spécifique des génomes de souches d'intérêt par **métagénomique**

Défis pour Migale

Projet de **très grande ampleur** (> 15 milliards de reads, plusieurs milliers d'échantillons)
Fonctionnement en mode **cohorte**
Benchmark, développement d'outils et automatisation
Étroite collaboration avec les experts (fiches descriptives, informations pertinentes...)
Rendre l'information exploitable et compréhensible au **grand public**

MetaPDO-Cheese : restitution – rapport web

Traçabilité / reproductibilité des analyses

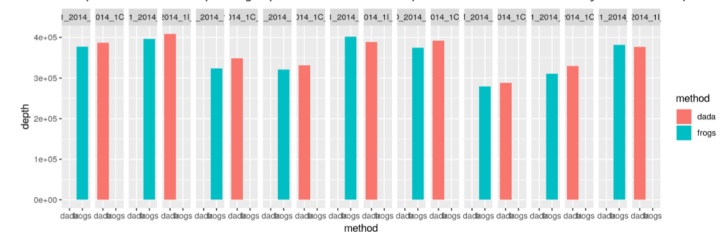
Rapports complets d'analyse sous forme de page web :

- Rmarkdown, Distill, KnitR , Gitlab Pages
- Détail des workflows /outils / versions / paramètres

1 Setup
2 Taxonomic composition
2.1 Sequencing depths
2.2 Rarefaction curves
2.3 Short taxa
2.4 Composition
2.5 beta-diversity
2.6 Adonis test

2.1 Sequencing depths

We first compare the effective sequencing depths of the two methods (to assess which one throws away the most reads)

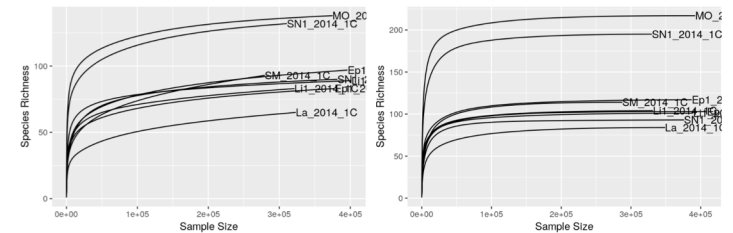


There is no difference in terms of effective depths between the two methods:

```
## Analysis of Variance Table
##
## Response: depth
##          Df Sum Sq Mean Sq F value Pr(>F)
## method  1 3.8947e+08 389465444  0.2265 0.6406
## Residuals 16 2.7514e+10 1719611286
```

2.2 Rarefaction curves

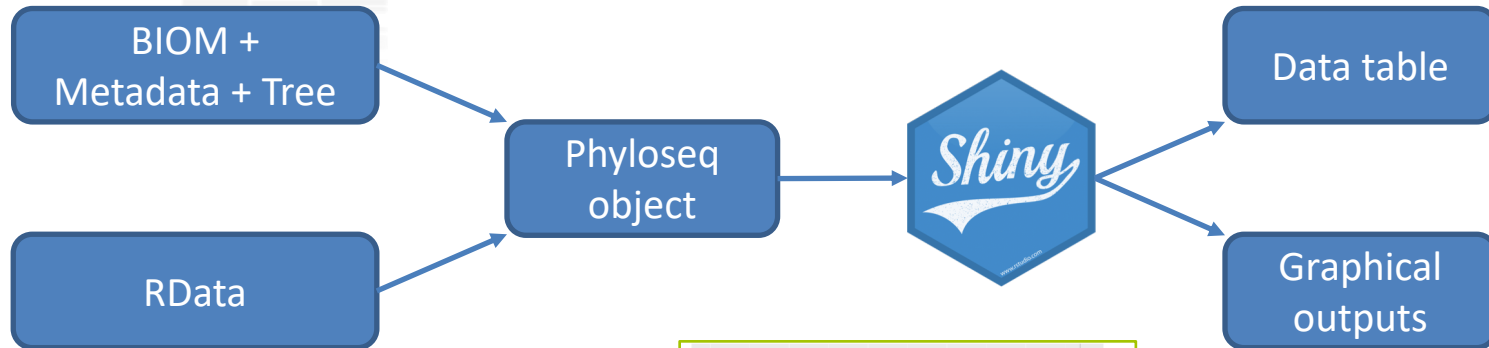
There are more taxa coming out of dada than of frogs but rarefaction curves seem to plateau in both cases:



GitLab



MetaPDO-Cheese : restitution – Easy16S



Easy16S

Select dataset: Demo: Chaillou et al., 2015

Summary Global barplot Filtered barplot Heatmap Rarefaction curves α -diversity β -diversity MultiDimensional Scaling PCA

Phylogenetic tree Clustering Help

phyloseq-class experiment-level object
 otu_table() OTU Table: [508 taxa and 64 samples]
 sample_data() Sample Data: [64 samples by 3 sample variables]
 tax_table() Taxonomy Table: [508 taxa by 7 taxonomic ranks]
 phy_tree() Phylogenetic Tree: [508 tips and 567 internal nodes]

Class of sample_data

Tables

otu_table	tax_table	sample_data	aggregate_taxa							
Show 10 entries	Column visibility	Download								
OTU	DUT1.LOT08	DUT1.LOT05	DUT1.LOT03	DUT1.LOT07	DUT1.LOT06	DUT1.LOT01	DUT1.LOT04	DUT1.LOT10	MYT1.LOT05	MY
	AB	AB	AB	AB	AB	AB	AB	AB	AB	
otu_00520	0	0	0	0	0	0	0	0	0	0
otu_00555	4	0	8	3	15	4	2	0	0	12
otu_00568	0	0	0	0	0	0	0	0	0	0
otu_00566	0	3	0	0	0	0	0	0	0	0
otu_00569	12	3	26	16	38	0	4	0	0	19
otu_00545	0	0	0	0	0	0	0	0	0	0
otu_00527	0	0	0	0	0	0	0	0	0	0
otu_00521	0	0	0	0	0	0	0	0	0	0
otu_00526	0	0	0	0	0	0	0	0	0	0
otu_01028	0	0	0	0	0	0	0	0	0	0

Showing 1 to 10 of 508 entries

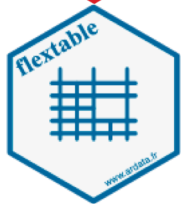


MetaPDO-Cheese : restitution - plaquette

Résultats accessibles à un public non expert:

Production automatique de fiches synthétiques (3 pages) à destination des représentants des filières

- Template au format docx
- KnitR , Officer, flextable



Fromages de Terroirs

Filière XXX

Des premiers résultats du projet **MetaPDOcheese** pour votre filière :
Les résultats bactériens sur les fromages

Rappel des échantillons envoyés :

- 10 échantillons de lait de cuve, dont XX en période hivernale et XX en période estivale.
- 3 échantillons de fromages pour chacune des fabrications correspondant aux 10 laits prélevés.
- Croûtes et pâtes ont été séparées et analysées séparément.

Pour votre filière, il y a donc au total : **70 échantillons**.

Sur chacun des échantillons, 2 analyses **métabarcoding** sont réalisées :

- L'une pour les Bactéries
- L'autre pour les levures / moisissures, aussi appelées **Fongi**.

Cette fiche vous présente les résultats obtenus sur les fromages (croûte et pâte) pour les Bactéries.

Diversité bactérienne présente par échantillon :

Le tableau ci-dessous renseigne les niveaux moyens de flores bactériennes pour les pâtes et les croûtes, ainsi que le nombre moyen d'OTUS recensé par échantillon et le nombre d'OTUS minoritaires (représentant moins de 5 %)

	Pâte			Croûte			Ttes filières	
	Médiane	Min	Max	Médiane	Médiane	Min	Max	Médiane
Niveau Flore 1 (u _{fc})								
Niveau Flore 2 (u _{fc})								
Niveau Flore 3 (u _{fc})								
Nombre d'OTUS total								
Nombre OTUS minoritaires								

Flore 1 : ; Flore 2 : ; Flore 3 :

L'OTU ou xxx correspond au plus petit niveau auquel la bactérie a pu être assignée : il peut donc s'agir du niveau famille, du niveau genre et dans le meilleur des cas, du niveau espèce.

Taxonomie assignée 1 à 20
L'axe des ordonnées donne les abondances

OTUS assignés

Taxonomie assignée 1 à 20
L'axe des ordonnées donne les abondances

OTUS assignés

Les histogrammes de gauche représentent l'abondance relative dans la pâte pour les 3 échantillons d'une même fabrication, ceux de droite, la même chose sur la croûte. Il est impossible de représenter l'ensemble des OTUS : la partie blanche (ou noire) du haut des histogrammes représente la proportion occupée par les OTUS non énumérés. Les principaux OTUS présents dans la pâte ou la croûte n'étant pas les mêmes, les légendes sont différentes pour les pâtes et les croûtes, ainsi que les couleurs.

Les 60 histogrammes disponibles sont présentés en page 3 et 4. Que ce soit pour la pâte ou la croûte, les échantillons issus d'une même fabrication sont généralement très proches, ce qui illustre l'homogénéité des lots.



Conclusion

- Les projets de recherche permettent de **calibrer/mettre à jour les pipelines** qui pourront ensuite être proposés :
 - en service d'analyse (**collaboration**)
 - En **module de formation** , avec possibilité **d'accompagnement** post-formation
- L'effort est mis sur **la traçabilité des analyses**, la production de **rapports complets et conviviaux**, exploitables par les collaborateurs
- Le triptyque **formation / accompagnement / mise à disposition des protocoles** d'analyses bioinfos et stats permet **d'autonomiser** les collaborateurs qui le souhaitent
- La **formalisation** des projets (besoins / projets / livrables / délais) permet de traiter un grand nombre de projets avec des ressources limitées.

Sophie SCHBATH

Responsable scientifique

Valentin LOUX

Responsable opérationnel

Véronique MARTIN

Formation, Outils et Banques de données

Valérie VIDAL

Bases de données

Sandra DEROZIER

Conception et développement

Olivier RUÉ

Analyse de données

Cédric MIDOUX (IRSTEA)

Analyse de données

Olivier Inizan

Implémentation et déploiement automatique de pipelines

Mahendra Mariadassou

Biostats

Hélène Chiapello

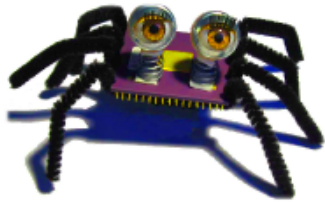
Formation

Sam AH-LONE

Apprentie Master2 – Projet Mosaic

David Christiany

Intégration d'outils sous Galaxy - IFB



help.migale@jouy.inra.fr

<http://migale.jouy.inra.fr>

