

Présentation de l'avancement du projet

Florian CHARRIAT

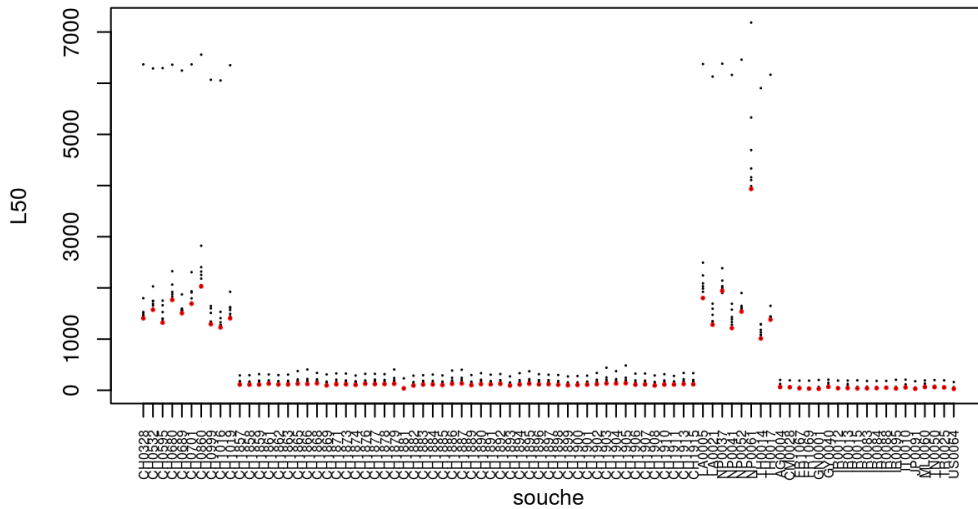
03 Avril 2018

Assemblage (ABYSS)

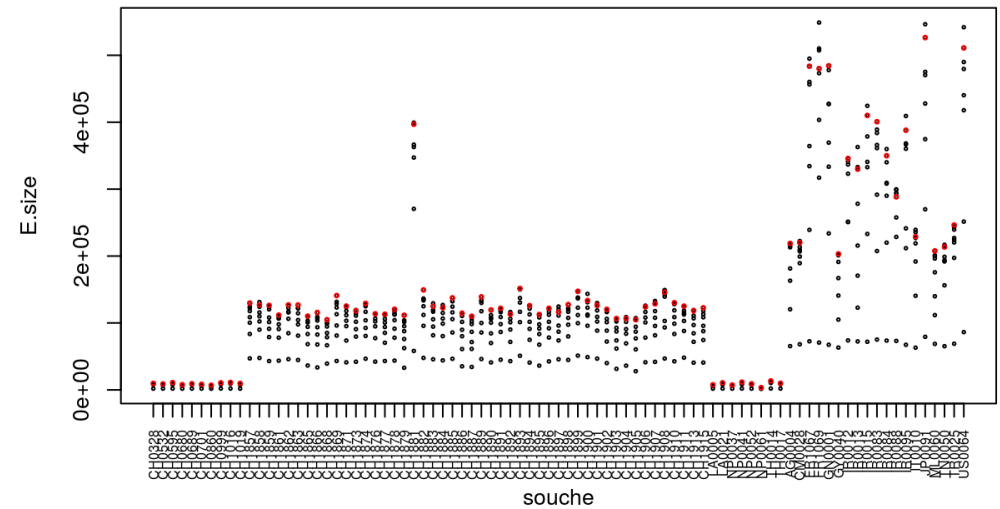
- 85 souches assemblées
 - 66 souches des terrasses
 - 19 souches séquencées par Adel
- Différentes tailles de k-mère utilisées pour chaque souche
 - Visualisation des statistiques de qualité pour choisir le meilleur assemblage pour chaque souche
 - Sélection par N50
 - Vérification de la sélection par visualisation des autres statistiques
- Visualisation de la qualité des 85 assemblages sélectionnés
 - Visualiser la qualité moyenne des assemblages
 - Filtrer les assemblages de mauvaise qualité

Sélection du meilleur assemblage

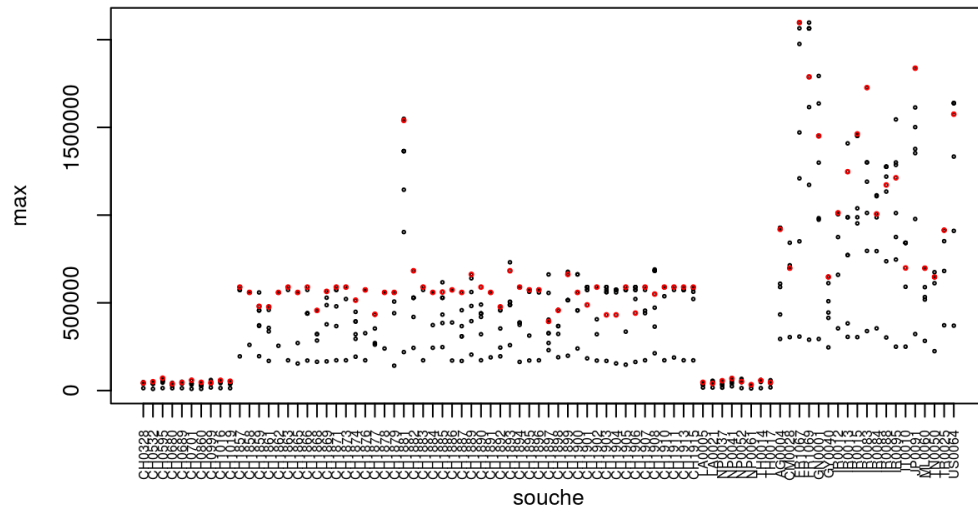
Graphique n°1 : Assemblage de chaque souche en fonction du L50



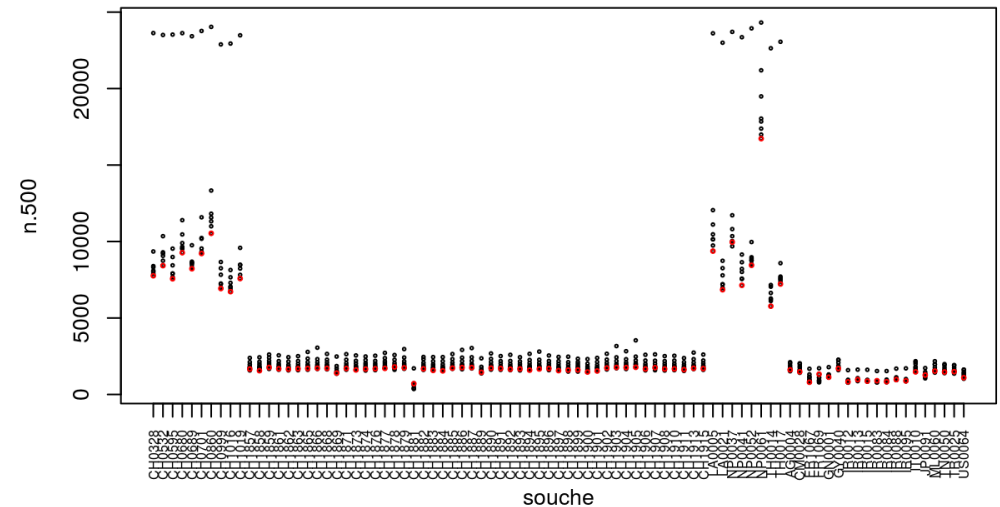
Graphique n°2 : Assemblage de chaque souche en fonction de l'E-size



Graphique n°3 : Assemblage de chaque souche en fonction de la longueur max des contigs

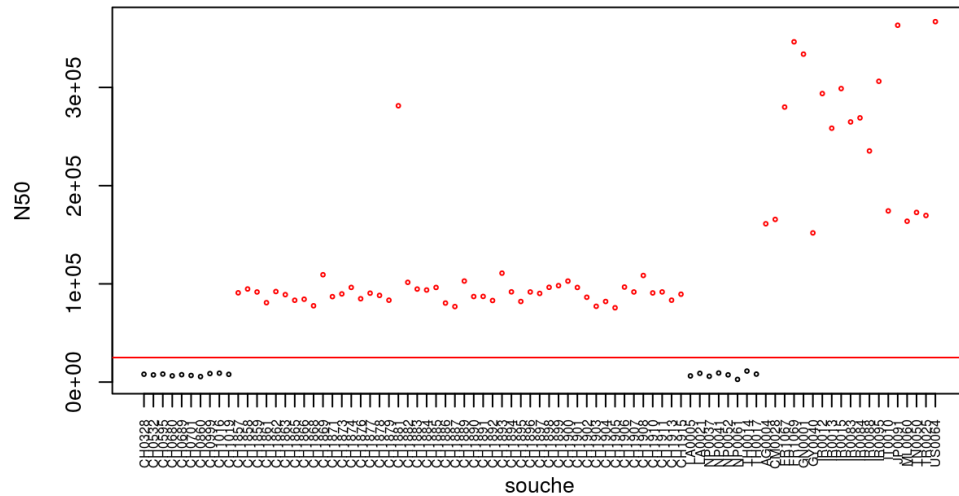


Graphique n°4 : Assemblage de chaque souche en fonction du nombre de contigs supérieur à 500

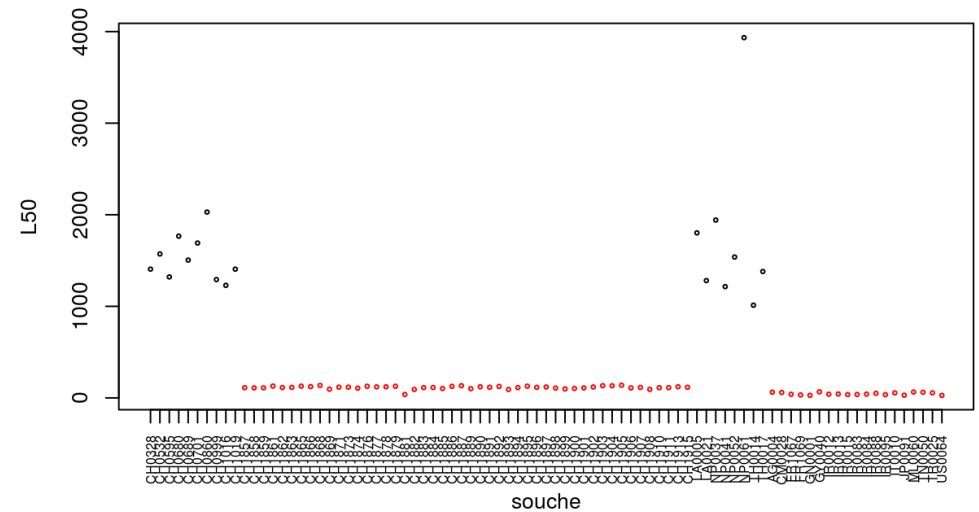


Qualité des assemblages sélectionnés

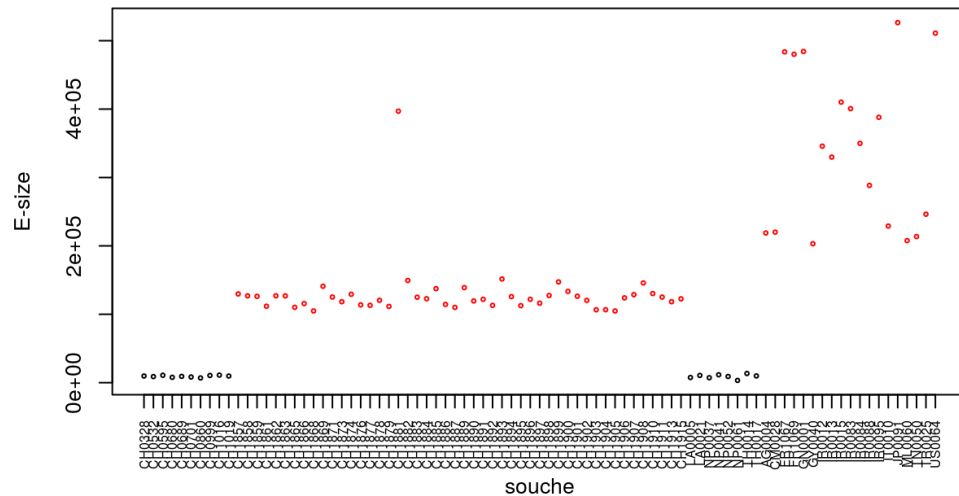
Graphique n°5 : Assemblage de chaque souche en fonction du N50



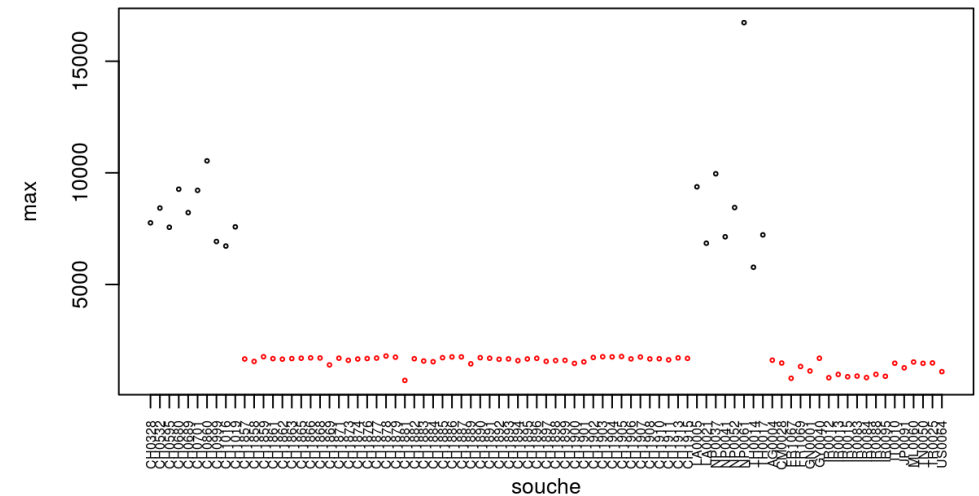
Graphique n°6 : Assemblage de chaque souche en fonction du L50



Graphique n°7 : Assemblage de chaque souche en fonction de l'E-size



Graphique n°8 : Assemblage de chaque souche en fonction du nombre de contigs supérieur à 500



repeatMasker

- Masquer les éléments répétés des 67 souches sélectionnées.
- Pour chaque souche :
 - 4.67 à 8.18 % des séquences ont été masquées par repeatMasker
 - Dans les assemblages certains scaffolds ne sont constitués que d'éléments répétés
 - Mise en place d'un script qui élimine ces scaffolds

Annotation (BRAKER)

- 154 souches à annoter :
 - 67 souches assemblées précédemment
 - 87 souches de la littérature
 - Plus des souches en cours de séquençage
- Pour annoter :
 - Génome à annoter
 - Choix de l'assemblage entre ceux effectués avec ABYSS et ceux réalisés par Sebastien
 - Données RNAseq alignées avec le génome à annoter
 - topHat : Aligner toutes les données RNAseq sélectionnées
 - Samtools : Merger les alignements de chaque donnée RNAseq

Mapping RNAseq

- 32 données RNAseq sélectionnées
 - 2 données provenant du projet GEMO
 - 15 données de GY11
 - 3 données de Magnaburkho
 - 12 données InPlanta
- Traitement des données Inplanta :
 - Mapping sur le génome du riz pour récupérer seulement les reads de Magnaporthe oryzae
- Mise en forme des assemblages pour l'alignement :
 - Elimination des scaffolds de moins de 500 pb.
 - Les identifiants ont été modifié :
 - Les scaffolds seront numérotés en fonction de leur longueur.
 - En description de chaque scaffold sa longueur sera donnée.