



Introduction atelier variant

Maria Bernard - INRA de Jouy en Josas

École de bioinformatique AVIESAN-IFB 2017



CNRS UPMC
Station Biologique
Roscoff

Atelier variant - objectifs

Processus d'analyse de données de séquences, du filtre qualité à la détection de variant:

- SNP et petits indel grâce en particulier à la suite GATK
 - Intervenants : Manuel, Guillaume, Kenza
- Variations structurales en particulier les grandes délétions grâce à deux outils.
 - Intervenants : Guillaume
- Utilisation de R pour visualiser quelques statistiques
 - Intervenante : Elodie

Atelier variant - le cluster de l'IGR

L'institut Gustave Roussy, met à notre disposition un petit cluster de calcul de 3 noeuds

Your turn! Se connecter au cluster

Sous Windows avec MobaXterm

Session : ssh

Host : galaxy.gustaveroussy.fr

Specify username : coché et complété

Advanced SSH settings : X11-Forwarding

Sous Mac/Linux dans un terminal

```
ssh -XY USERNAME@gustaveroussy.fr
```

Atelier variant - le cluster de l'IGR

Vous avez vu comment se connecter à un noeud particulier d'un cluster.

Tout le monde est au départ connecté sur le même noeud. Deux autres noeuds sont disponibles : node01 et node02. Ici le cluster est sous ordonnanceur slurm.

Your turn ! Répartissons nous sur les 3 noeuds disponibles

```
# 1/3 reste sur le noeud de connexion, 1/3 sur node01, 1/3 sur node02
```

```
srun -w NODE_NAME --pty bash
```

Atelier variant - jeux de données 1 : SNP / indels

Depuis que l'homme fait de l'élevage, il essaie de faire en sorte de toujours améliorer sa production, que ce soit en quantité ou en qualité. Les technologies de génotypage permettent maintenant de sélectionner les mâles reproducteurs en fonction du fond génétique qu'ils vont pouvoir transmettre à leur descendance.

Chez le bovin, on sait qu'il y a un QTL lié à la production de lait, sur le chromosome 6 plus exactement sur une région de 700 kb. On retrouve dans cette région 7 gènes.



Atelier variant - jeux de données 1 : SNP / indels

Les échantillons QTL+ sont caractérisés par une diminution de la production en lait et une augmentation des concentrations en protéine et lipide.

Vous aurez à votre disposition :

- Un extrait des données de séquences d'un échantillon du projet 1000 génomes bovins, phénotypé comme QTL- : SRR1262731
- Le fichier gvcf d'un second échantillon phénotypé QTL+ : SRR1205992

Your turn ! Quelle mutation est responsable de ce QTL ?

Atelier variant - jeux de données 2 : variations structurales

Zymoseptoria tritici

Champignon ascomycète, pathogène du blé tendre, responsable d'une maladie foliaire (septoriose).

Principale maladie du blé(jusqu'à 50% de perte de rendement).

Haploïde, Génome de 40 Mb séquencé en 2011.

13 chromosomes essentiels + 8 chromosomes accessoires

-> Exercice de détection des délétions sur une souche Française séquencée en Illumina .

