

## TP QTLMap

**Objectif : faire des analyses LD, LA et LDLA sur les 781 descendants (+ ascendants) du fichier simulé.**

Copier le répertoire du Cours dans votre propre espace :

```
cp -r /travail/COURS_RT/COURS_RT/TP_QTLMap /chezvous/TP_QTLMap
```

### 1. Les fichiers d'entrée dans QTLMap

Le répertoire « Fichiers » dans TP\_QTLMap contient les 4 fichiers de données :

- « carte » donne la position des marqueurs sur la carte génétique ;
- « geno\_garde » donne les génotypes aux marqueurs ;
- « pedig » donne la généalogie ;
- « pheno » donne les phénotypes.

Nous décrivons oralement les formats de ces fichiers en TP et vous pourrez trouver plus d'info dans la doc de QTLMap si besoin (<http://www.inra.fr/qtlmap>).

Le répertoire « Fichiers » contient également le fichier « model » qui décrit les caractères à analyser.

Enfin, pour lancer QTLMap, il faut préparer un fichier décrivant à l'aide de différents paramètres l'analyse à réaliser. Un exemple est disponible dans « Fichiers » avec le fichier « param\_C1 » qu'il vous faut compléter. Nous analyserons le chromosome 1 en détail en TP, mais vous pouvez aussi analyser le 2 si vous voulez.

Ceci étant fait, vous avez tous les fichiers nécessaires pour exécuter QTLMap.

### 2. Lancement d'une analyse LD

Dans le répertoire « Fichiers », vous avez un script « LD.sh » pour lancer QTLMap :

```
#!/bin/bash
#$ -q rapide
#$ -l hvmem=10G
#
qtlmap param_C1 --snp --calcul=26 --hdam --qtl=1
```

L'exécution de QTLMap est demandée dans la dernière ligne de ce script :

```
qtlmap <nom du fichier paramètres> <options>
```

Dans notre exemple, le fichier paramètre s'appelle « param\_C1 » et les options sont :

- « --snp » pour phaser les parents avec la méthode de Favier *et al.* (2009) ;
- « --calcul=26 » pour faire une analyse LD
- « --hdam » pour prendre en compte les haplotypes maternels ;
- « --qtl=1 » pour faire une analyse 0 QTL vs 1 QTL.

Pour lancer l'analyse :

```
Taper qsub LD.sh
```

Vous pouvez suivre l'exécution de votre travail en tapant « qstat ». Vous devriez voir rapidement votre job se terminer avec un code erreur. Pour comprendre ce qui s'est passé, vous devez aller voir le fichier « LD.sh.o<numéro de votre job> ». Vous verrez que le fichier de génotypages contient une incompatibilité entre parents et descendants. Vous pouvez remarquer aussi que depuis 2 jours vous avez fait vos analyses sans souci !

Différentes solutions sont possibles pour régler le problème : enlever le génotype du descendant pour le marqueur à problème (un peu compliqué à faire rapidement) ; enlever le descendant du fichier généalogie (un peu violent) ; enlever le marqueur de l'analyse en mettant un « 0 » dans la dernière colonne du fichier carte. Les fichiers préparés pour ce TP l'ont été en prenant la dernière option.

Vous pouvez relancer l'analyse une fois l'erreur enlevée. Normalement, cela devrait marcher cette fois. En fin de travail, vous devriez avoir obtenu différents fichiers de résultats (en fonction de ce que vous aurez mis dans votre fichier paramètres). Pour notre exemple, dans :

<out\_output> vous aurez un listing avec plein d'informations

<out\_phases> vous aurez les phases des pères

<out\_lrtsires> vous aurez les LRT obtenus à chaque position

<out\_pded> vous aurez les probabilités de transmissions pour chaque descendant à chaque position, père-descendant et mère-descendant

<out\_pdedjoin> la même chose en probabilité conjointes {père-mère}-descendant

<out\_pateff> vous aurez les contributions au LRT de chaque famille de père à chaque position

<out\_summary> vous aurez un résumé des résultats par famille de père.

Nous commenterons ces différents fichiers oralement pendant le TP.

Pour savoir si vous rejetez l'hypothèse d'absence d'un QTL sur le chromosome 1, vous devez avoir les seuils de rejet de cette hypothèse. Pour cela, il faut faire un grand nombre de simulations sous H0 (au moins 1 000).

Pour le TP, vous allez simplement en lancer 2 (!) en ajoutant l'option « --nsim=2 » à votre ligne de commande « qtlmap ... ». Il faut changer votre fichier paramètres avant de lancer cette exécution :

- Ajouter un nom pour le fichier de stockage des LRT simulés (out\_maxlrt=<nom>)
- Changer les noms de vos fichiers de sortie (les out\_...) sinon vous allez écraser vos beaux résultats durement obtenus.
- Une ruse parfois intéressante consiste à diminuer le pas d'exploration du chromosome (par exemple de 1 à 5 cM) pour réaliser des simulations, surtout sous H0 (<opt\_step>).

Vous trouverez le résultat obtenu avec 1000 simulations dans le fichier /TPQTLMap/Chromosome1/LD/Seuils.

Au bout du compte, vous devriez conclure qu'il y a un QTL significatif à 10% à la position 0,2 sur le chromosome 1. Vous avez les sorties attendues dans le répertoire /TPQTLMap/Chromosome1/LD.

Si vous avez fait le chromosome 2, les sorties et les seuils sur 1 000 simulations sous  $H_0$ , sont disponibles dans /TPQTLMap/Chromosome2/LD.

Vous pouvez comparer ces résultats avec ce que vous aviez obtenu avec l'analyse GWAS ou avec GS3 qui, sur le principe, utilisent la même information (le DL).

### **3. Lancement d'une analyse LA**

Vous pouvez refaire la même chose en demandant une analyse de liaison (LA) avec l'option « --calcul=4 » et observer les différences. Attention : pensez bien à changer les noms de vos fichiers de sortie !

Vous trouverez les résultats et les seuils dans /TPQTLMap/Chromosome1/LA.

Normalement, vous devez conclure à l'absence de QTL sur le chromosome 1.

### **4. Lancement d'une analyse LDLA**

Enfin, vous pouvez faire une analyse LDLA avec l'option « --calcul=28 ».

Vous trouverez les résultats et les seuils dans /TPQTLMap/Chromosome1/LDLA.

Normalement, vous devez conclure à la présence d'un QTL sur le chromosome 1 ( $p=3\%$ ) ... à la position 0,76.

### **5. Analyse d'un sous fichier**

Le fichier /TPQTLMap/Fichiers/pedigvieu contient le pedigree des familles les plus anciennes du pedigree : 338 descendants issus de 7 pères. Reprenez l'analyse LDLA en considérant uniquement ces familles. Pour cela, vous pouvez ré utiliser vos autres fichiers de données, sans les modifier, car QTLMap se base sur le fichier généalogie pour définir l'échantillon à analyser. Vous aurez simplement un message dans le listing disant que certains animaux du fichier de performances ne sont pas utilisés. Voyez la puissance obtenue avec ce sous fichier qui pourrait correspondre à une première expérimentation ... qui demande à être complétée.

Les résultats et seuils sont dans /TPQTLMap/Gener1.