

# Contrôle de l'erreur de première espèce. Méthode Muller et al

Müller et al (2011)<sup>1</sup> ont proposé une méthode pour le contrôle de l'erreur de première espèce dans les analyses d'association ou de liaison, basée sur une méthode rapide de simulation. Cette méthode est applicable aux approches de type « genome scan » dans lesquels chaque chromosome est exploré en testant successivement la présence d'un QTL en une succession de localisations. La méthode prend en compte la non-indépendance, engendrée par la liaison génétique, entre les distributions des statistiques de test aux différentes positions. Elle le fait très rapidement par le calcul des éléments de la matrice des covariances entre ces statistiques de test et sa décomposition spectrale. Le logiciel développé dans le cadre de Rules & Tools est une application directe de l'algorithme proposé. Il permet d'obtenir sans difficulté des seuils de rejets objectifs de l'hypothèse d'absence de QTL sur un groupe de liaison.

## 1 Format des fichiers utilisateurs

Le programme MULLER utilise un fichier de génotype au format AIPL. Les identifiants des animaux sont numérique.

### 1.1 Le fichier des génotypes

Le fichier des données de génotype contient l'information au marqueur pour chacun des descendants. Ce format renseigne l'identifiant de l'individu en première colonne (numérique) et en deuxième colonne une chaîne de caractère décrivant l'information moléculaire. Ce fichier ne contient pas d'entête

#### 1.1.1 Le format AIPL

Le génotype d'un individu est codé sur une chaîne de caractère. Le codage utilise 4 valeurs entières. Une valeur est associée à chaque marqueur SNP :

- 0 pour le statut homozygote au premier allèle ;
- 1 le statut hétérozygote ;
- 2 le statut homozygote au deuxième allèle ;
- 5 le statut "valeur manquante".

```
4500 11211211
4501 01112102
4502 01211102
4503 01111102
...
```

---

1. Müller BU, Stich B, Piepho HP, 2011. A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. *Heredity* 106 : 825-831

## 1.2 Le fichier des animaux phénotypés

Ce fichier contient une unique colonne avec les identifiants des animaux phénotypés utilisés dans blupf90

```
4502
4503
..
```

## 2 Exécution

### 2.1 Arguments en entrée du programme

- --geno <pathfile>, -g <pathfile> : Le fichier des génotypes ;
- --matfile <patternfile>, -m <patternfile> : Le masque des fichiers en sorties de blupf90 (mat\_snpreg) ;
- --animal <pathfile>, -a <pathfile> : Le fichier contenant les animaux utilisés dans le calcul des blupf90 (animaux phénotypés) ;
- --npos <integer>, -n <integer> : Le nombre de positions testés ;
- --vare <float>, -v <float> : La variance résiduelle utilisée en paramètre des blupf90 ;
- --nlevel <integer>, -l <integer> : Le cumul du nombre de niveau des effets fixés du modèle utilisés par blupf90 ;
- --nsim <integer>, -s <integer> : Le nombre de simulation pour l'obtention des seuils de rejets.

#### 2.1.1 Arguments optionnels

- --idlength <integer>, -i <integer> : nombre de caractère utilise pour le codage de l'identifiant numérique dans le fichier de génotypage (10 par défaut).

### 2.2 Exemple

Execution dans un répertoire contenant le fichier de génotype genotypeb et les fichiers mat\_snpreg1 à mat\_snpreg10. Le calcul est sur 10 positions testées (10 SNP) avec une variance résiduelle de 20.831895. Le nombre de niveau cumulé de l'ensemble des effets fixés du modèle est égale à 15.

```
muller -g ./genotypeb -m ./mat_snpreg -n 10 -v 20.831895 -l 15 -a animal_list -s 1000
```