Annexe Mémoire

Capucine

Table of Contents

### Notations :

* : fréquence de l’allèle majeur (l’allèle le plus représenté dans la population étudiée)
* : fréquence de l’allèle mineur (le moins représenté)
* Equilibre de Hardy-Weinberg (HWE)

### Pour 1 croisement :

#### 1 - Variance du score d’un marqueur dans une population F2 sous HWE

Un marqueur est supposé être bi-allélique, son score (= nombre de copie de l’allèle mineur) est défini comme suit :

La variance du marqueur peut être calculée comme pour n’importe quelle variable en utilisant :

D’où :

#### 2 - Variance d’un QTL

Le marqueur suivant est supposé être dans une situation avec un effet additif :

Par souci de simplicité, il est supposé qu’il n’y a pas d’effet de dominance ().

L’espérance phénotypique pour un modèle simple locus est :

L’espérance de la valeur phénotypique au carré est de :

Par conséquent, la variance est :

A partir de l’expression précédente, il est possible de dériver une expression de : En supposant que , et suivant la précédente expression :

Mollandin et al., 2022 ont sans doute supposés que

#### 3 - Faire coïncider la variance simulée espérée (expected) et la

variance effectivement simulée (realized)

##### Formule de Mollandin et al., 2022 pour simuler 1 croisement :

Cependant, Mollandin et al, 2022 ont utilisé une autre formule pour simuler l’effet d’un SNP dans une population comportant un croisement, selon la proportion de variance génétique totale portée par ce QTL:

(1)

* : effet du SNP i
* : provient d’une distribution uniforme {-1,1} qui permet d’allouer une valeur positive ou négative à l’effet du SNP i
* : variance additive
* : proportion de variance additive ( ) expliquée par le SNP i
* ( ) : fréquence de l’allèle mineur (majeur) du SNP i ()

##### Adaptation de la formule de Mollandin et al., 2022 :

En supposant que la formule (1) est utilisée pour un seul QTL avec , on a :

(2)

Précedemment, l’expression de a été dérivée :

Les phénotypes simulés sont de la forme : avec la matrice de génotypage du marqueur, standardisée ou non, et l’effet du QTL. La variance du phénotype d’un QTL simulé est donc :

Or, pour les simulations, il est souhaitable que :

Dès lors, selon si est standardisée ou non : La variance du score du marqueur a été démontrée auparavant et est égale à

Cependant, on remarque que la variance réellement simulée n’est pas exactement égale à la variance utilisée pour simuler Ainsi, pour parvenir à faire coïncider variance simulée et variance obtenue, il faut changer le facteur qui devient alors lorsque X est standardisée, dans l’expression de .

En reprenant les calculs pour , il suit :

Pour X standardisée : , donc la variance effectivement simulée est bien égale à la variance théoriquement simulée.

La formule (1) devient donc :

(3)

### Plusieurs croisements :

#### 1 - Variance du score du marqueur dans une F2 sous HWE dans une population avec plusieurs croisements

##### Pour 2 croisements

Pour le croisement 1 :

Pour le croisement 2 :

La formule de la variance avec 2 croisements :

On calcule les éléments nécessaires :

Et,

Similairement pour le 2è croisement :

Et,

Puis,

Similairement pour le croisement 2 :

D’où :

Similairement pour le croisement 2 :

Ainsi, on remplace dans la formule générale de la variance :

On a donc :

On tente de réduire et simplifier :

Si on considère que et que , d’où , on a :

##### Généralisation

Pour

On veut démontrer que la variance du score du marqueur est bien toujours égale à , quelque soit le nombre de croisement (avec , et , d’où ).

Tout d’abord, on substitue les valeurs dans nos constantes:

Ensuite,

On réinjecte dans l’expression initale :

Première partie de la somme :

Deuxième partie de la somme :

On remplace dans l’expression initiale :

On a donc bien ce qu’on voulait démontrer, la variance du score des marqueurs est bien égale à quelque soit le nombre de croisement.

#### 2 - Variance génétique à un QTL avec plusieurs croisements

##### Pour 2 croisements :

Il est possible d’étendre la formule à une situation avec plusieurs croisements. Pour cela, commençons par une situation avec deux croisements par souci de simplicité.

On suppose que le marqueur suivant dépend uniquement d’effets additifs (donc l’hétérozygote correspond à la moitié de l’homozygote). On considère également que l’on se place dans une population NAM où le parent central (référence) a un allèle valant 0.

Croisement 1 :

Croisement 2 :

Dans ce cas, la formule de la variance conditionnelle pour une partition disjointe dans un espace probabiliste (croisement) est utilisée.

Dans un cas avec deux croisements, on a :

Où et sont les probabilités d’être dans le croisement 1 ou 2 respectivement, et .

Il faut estimer les composantes suivantes :

Development of the different expressions

Calculs intermédiaires :

Suite développement expressions :

Maintenant, on calcule les variances. Pour cela, on utilise :

D’où :

Et,

On remplace dans la formule de la variance :

D’où,

On simplifie en regroupant les termes par croisements et par effets () :

D’où :

On pose :

Ainsi,

Ensuite, on cherche à exprimer l’effet d’un QTL à partir de l’expression de sa variance (puisque qu’il n’y a pas de dominance :

###### Effet constant : b = c

ù D’où,

###### Effet en série allélique : c = sb

##### Pour 3 croisements :

On note l’effet du 3e croisement :

Croisement 1 :

Croisement 2 :

Croisement 3 :

Donc on peut calculer la variance avec ces éléments, puis simplifier :

Avec

###### Effet constant : b = c = d

On peut déduire :

###### Effet en série allélique : c = sb, d = s^2b

##### Généralisation pour nc croisements :

###### Effet constant

Pour

Avec,

D’où :

Le facteur 1/2, voir démonstration plus loin.

###### Effets en série allélique

, etc…

Pour

D’où :

#### 3 - Faire coïncider la variance simulée espérée (expected) et la variance effectivement simulée (realized) pour plusieurs croisements

##### Pour 2 croisements à effet constant :

Comme démontré plus haut, la variance théorique du score du marqueur est égale à .

En prenant les mêmes considérations : , et , d’où : La valeur du dénominateur est égale à :

Donc

Ainsi, en reprenant :

Or, si on considère que Dès lors, on a :

Il faut donc bien un facteur devant l’effet b.

##### Généralisation :

On suppose que et que , d’où

On veut démontrer que quelque soit le nombre de croisement, le dénominateur est bien égal à .

Tout d’abord, on substitue les valeurs dans nos constantes:

On réinjecte dans l’expression initale :

Première partie de la somme:

Deuxième partie :

En remplacant dans la somme :

On retombe donc bien sur ce qu’on voulait démontrer :

Ainsi, cela implique que l’on rajoute un facteur pour exprimer l’effet de .

Les effets égaux sont un cas particulier de la série allélique lorsque . Par extension, pour les effets en série allélique, ce facteur est toujours constant à , dans la mesure où seul le dénominateur change en fonction de la valeur de .