

14/02/24 - 28/06/24



Etude d'Association Pangénomique sur la locomotion chez *Caenorhabditis elegans*

Laboratoire d'évolution expérimentale

Alix Boutheroue-Desmarais

Encadrée par : Henrique Teotonio, François Mallard

Peut-on prédire l'évolution ?

Peut-on prédire l'évolution ?



CRISTINA MITTERMEIER EXPEDITION SEALEGACY

Peut-on prédire l'évolution ?

Phenotype

Peut-on prédire l'évolution ?

Phenotype

- Variabilité environnementale
- Plasticité

Peut-on prédire l'évolution ?

Phenotype

- Variabilité environnementale
- Plasticité

→ Evolution experimentale

Peut-on prédire l'évolution ?

Phenotype

- Variabilité environnementale
- Plasticité

 Evolution experimentale

Genotype

Peut-on prédire l'évolution ?

Phenotype

- Variabilité environnementale
 - Plasticité
- 
- Evolution experimentale

Genotype

Architecture génétique : organisation des effets génétiques qui contrôlent un phenotype

Peut-on prédire l'évolution ?

Phenotype

- Variabilité environnementale
 - Plasticité
- 
- Evolution experimentale

Genotype

Architecture génétique : organisation des effets génétiques qui contrôlent un phenotype

- Organismes pas collection de phenotypes indépendants

Peut-on prédire l'évolution ?

Phenotype

- Variabilité environnementale
 - Plasticité
- 
- Evolution experimentale

Genotype

Architecture génétique : organisation des effets génétiques qui contrôlent un phenotype

- Organismes pas collection de phenotypes indépendants
 - Corrélations génétique et environnementale

Peut-on prédire l'évolution ?

Phenotype

- Variabilité environnementale
 - Plasticité
- 
- Evolution experimentale

Genotype

Architecture génétique : organisation des effets génétiques qui contrôlent un phenotype

- Organismes pas collection de phenotypes indépendants
 - Corrélations génétique et environnementale

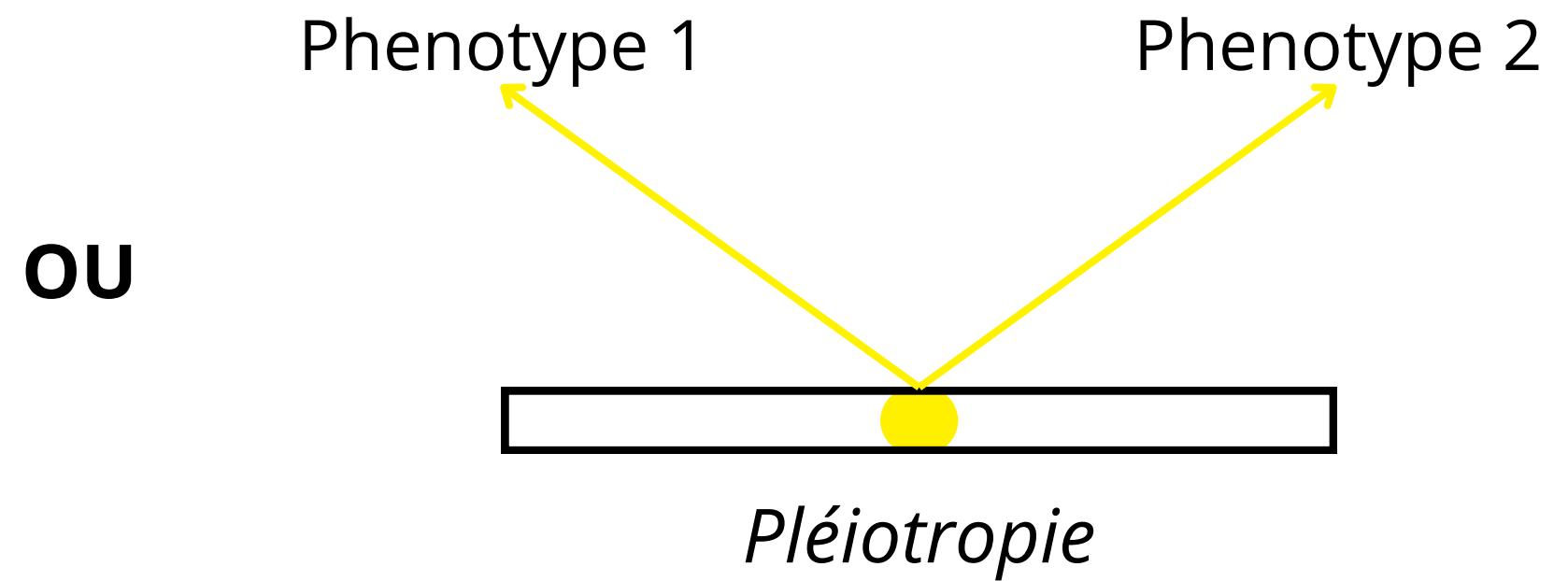
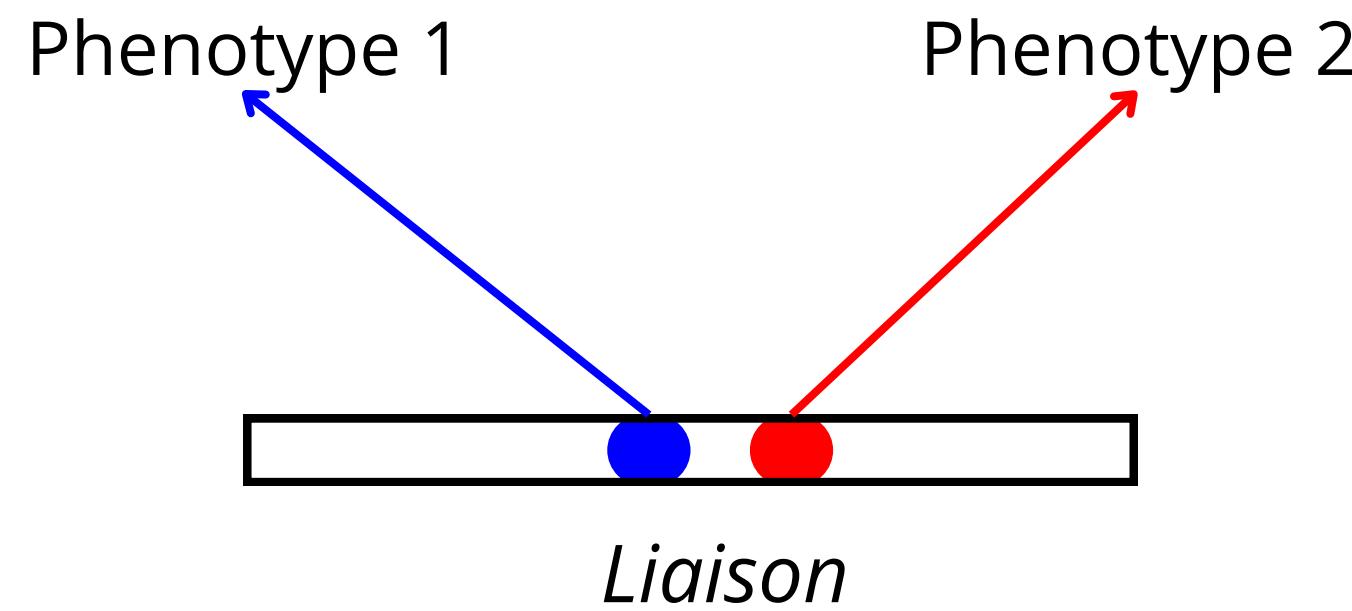


- Génome entier
- Multivariée

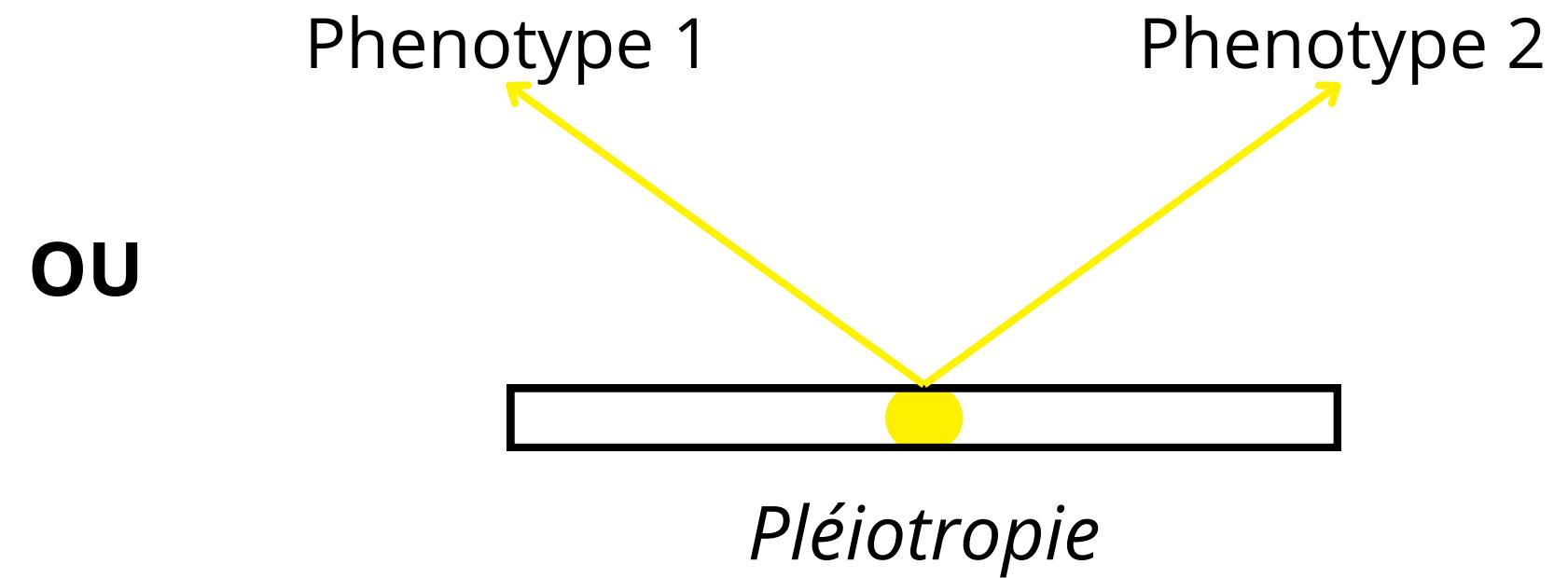
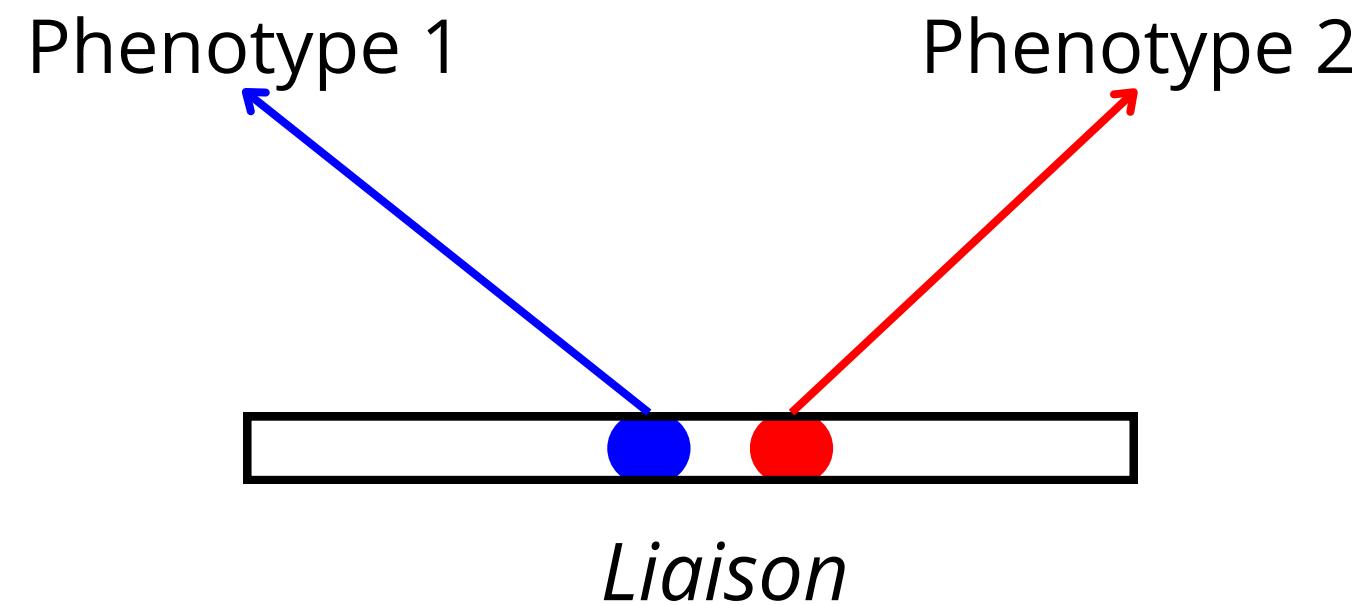
D'où viennent ces corrélations ?

D'où viennent ces corrélations ?
Au sein d'un environment ?

D'où viennent ces corrélations ?
Au sein d'un environment ?



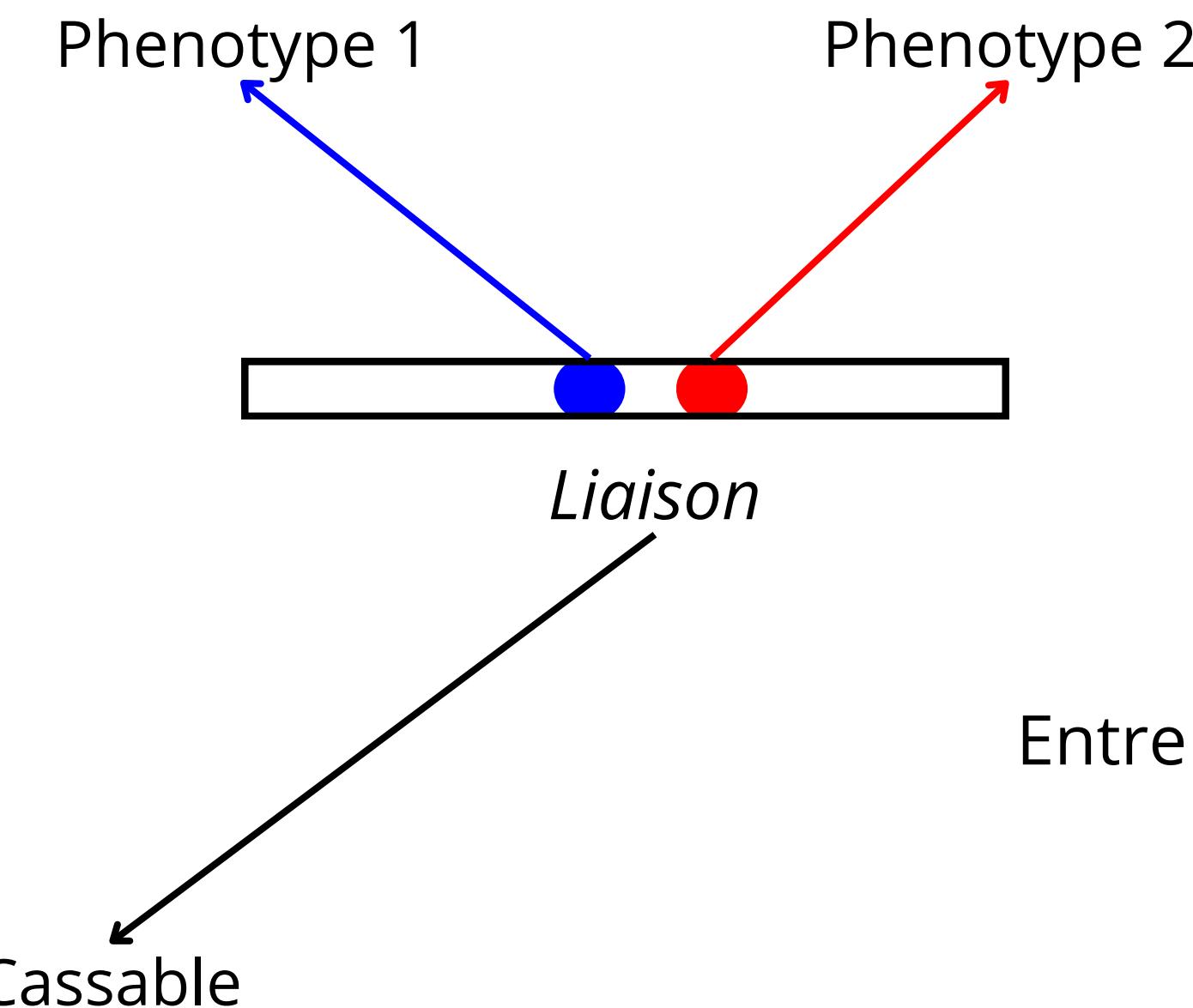
D'où viennent ces corrélations ?
Au sein d'un environment ?



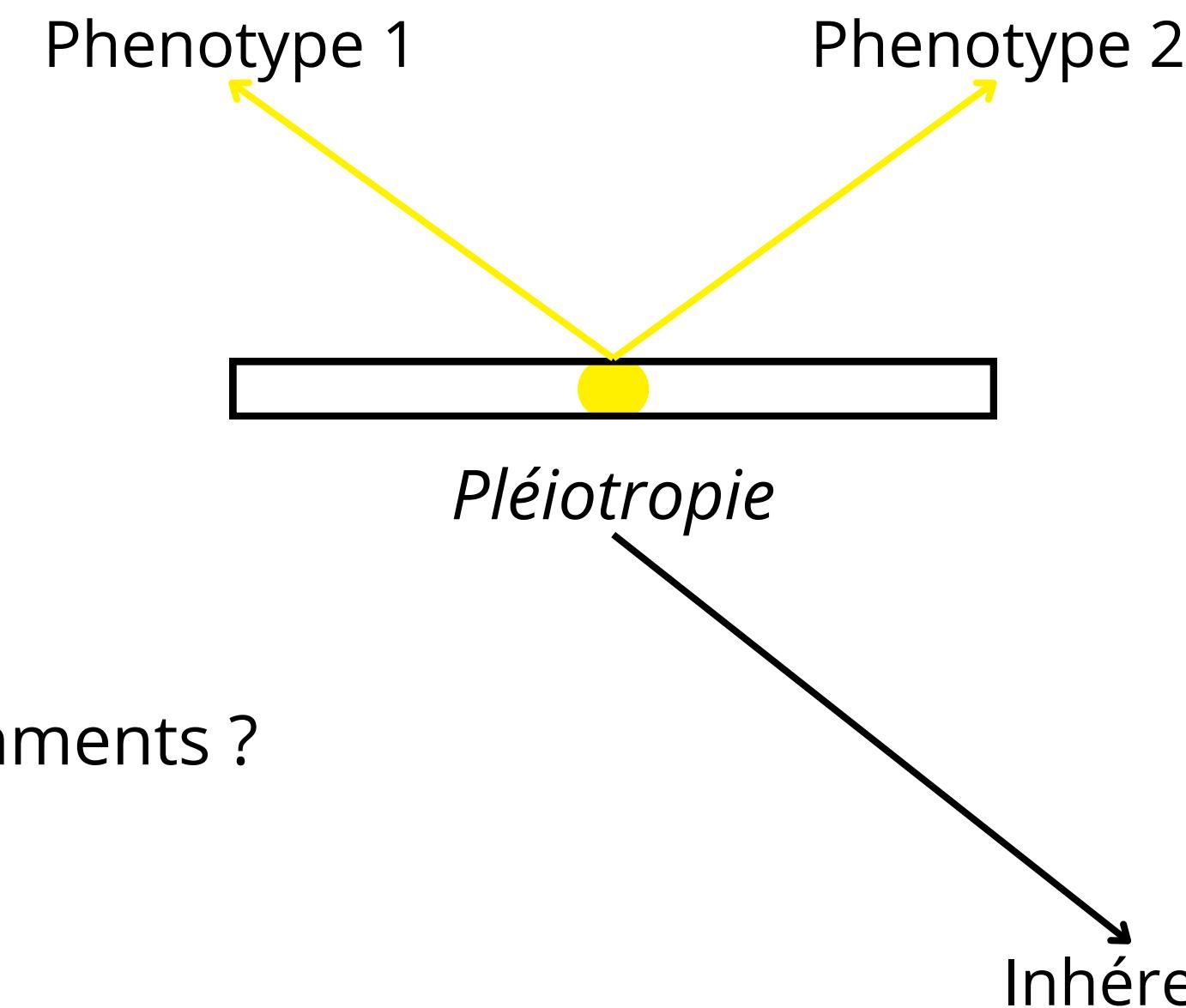
Entre les environments ?

D'où viennent ces corrélations ?

Au sein d'un environment ?



OU



Entre les environments ?

Explorer la structure polygénique de la locomotion dans deux environnements



Explorer la structure polygénique de la locomotion dans deux environnements

Genome Wide Association Study :

- Estimation des effets des variants génomiques
- Identification de variants statistiquement associés au phénotype



Explorer la structure polygénique de la locomotion dans deux environnements

Genome Wide Association Study :

- Estimation des effets des variants génomiques
- Identification de variants statistiquement associés au phénotype

Variant

=

Single Nucleotide Polymorphism (SNP)



Explorer la structure polygénique de la locomotion dans deux environnements

Genome Wide Association Study :

- Estimation des effets des variants génomiques
- Identification de variants statistiquement associés au phénotype

Variant

=

Single Nucleotide Polymorphism (SNP)

Approche classique

1 variant = 1 modèle



Explorer la structure polygénique de la locomotion dans deux environnements

Genome Wide Association Study :

- Estimation des effets des variants génomiques
- Identification de variants statistiquement associés au phénotype

Variant

=

Single Nucleotide Polymorphism (SNP)

Approche classique

GWAS by Genome
Best Unbiased
Linear Predictor
(GBLUP)

1 variant = 1 modèle



Explorer la structure polygénique de la locomotion dans deux environnements

Genome Wide Association Study :

- Estimation des effets des variants génomiques
- Identification de variants statistiquement associés au phénotype

Variant

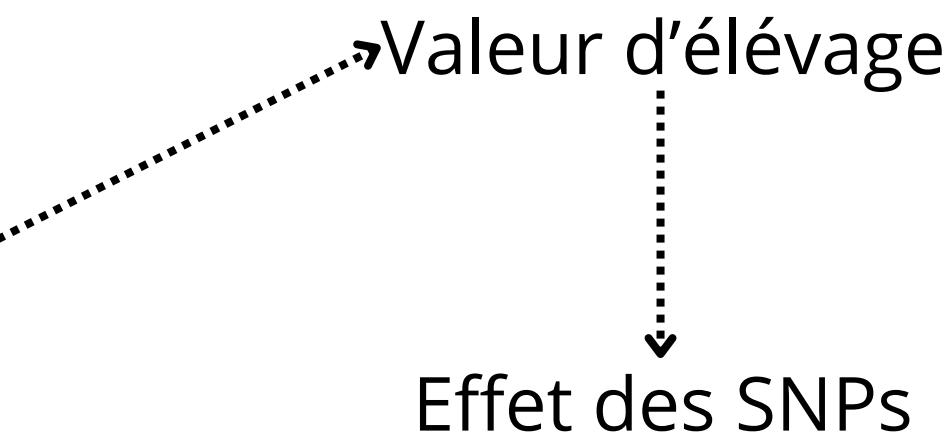
=

Single Nucleotide Polymorphism (SNP)

Approche classique

1 variant = 1 modèle

GWAS by Genome
Best Unbiased
Linear Predictor
(GBLUP)



Explorer la structure polygénique de la locomotion dans deux environnements

Genome Wide Association Study :

- Estimation des effets des variants génomiques
- Identification de variants statistiquement associés au phénotype

Variant

=

Single Nucleotide Polymorphism (SNP)

Approche classique

1 variant = 1 modèle

GWAS by Genome
Best Unbiased
Linear Predictor
(GBLUP)

Tous les variants =
1 modèle

Valeur d'élevage

Effet des SNPs



Explorer la structure polygénique de la locomotion dans deux environnements

Genome Wide Association Study :

- Estimation des effets des variants génomiques
- Identification de variants statistiquement associés au phénotype

Variant

=

Single Nucleotide Polymorphism (SNP)

Approche classique
1 variant = 1 modèle

GWAS by Genome
Best Unbiased
Linear Predictor
(GBLUP)

Tous les variants =
1 modèle

Approche
bayésienne

- Distribution à posteriori
- G-matrix



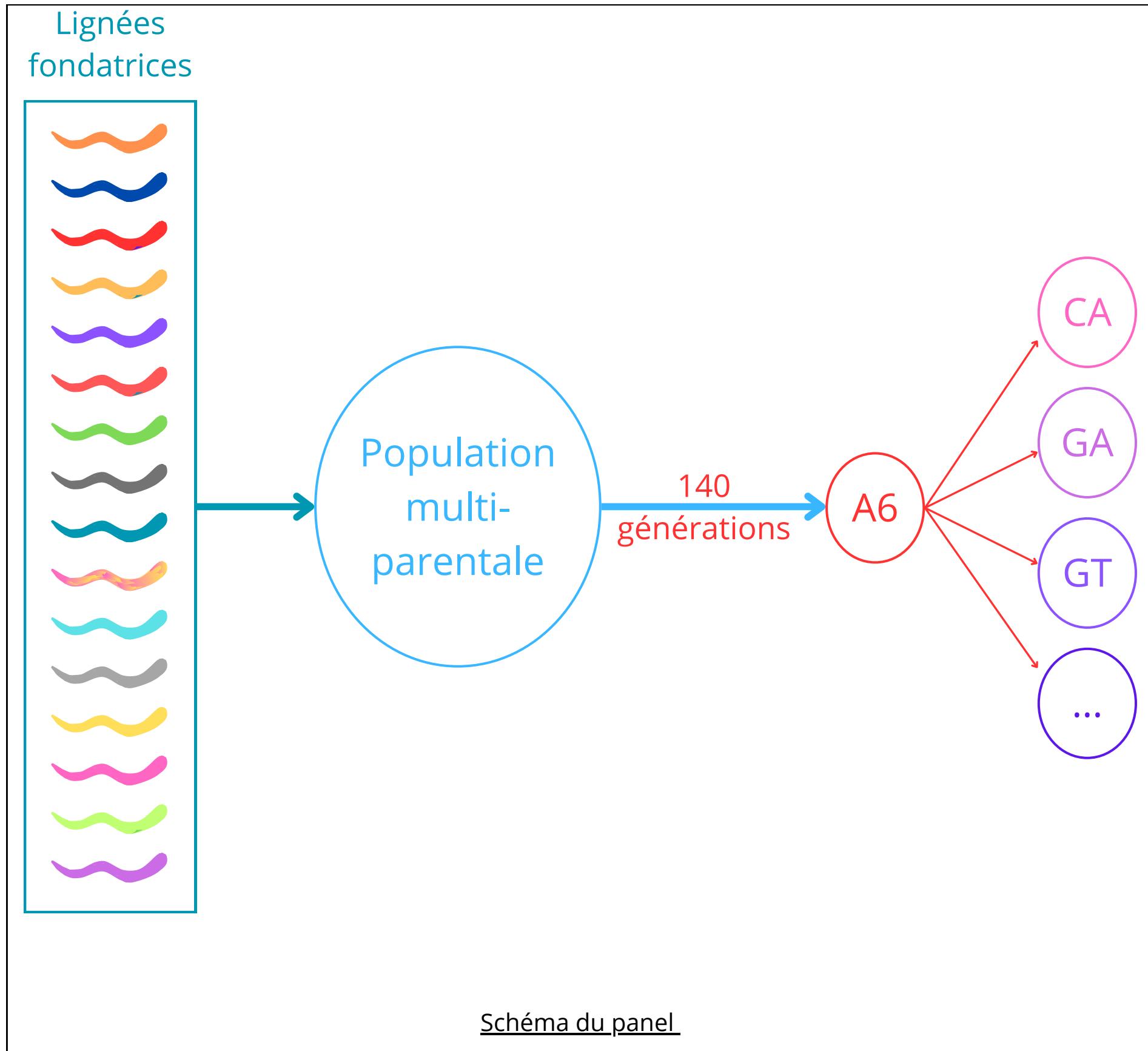
C. elegans

- Petite taille
- Cycle de vie court
- Cryopreservation
- Androdioecie
 - Male / Hermaphrodite
 - Hermaphrodite : auto-fécondation



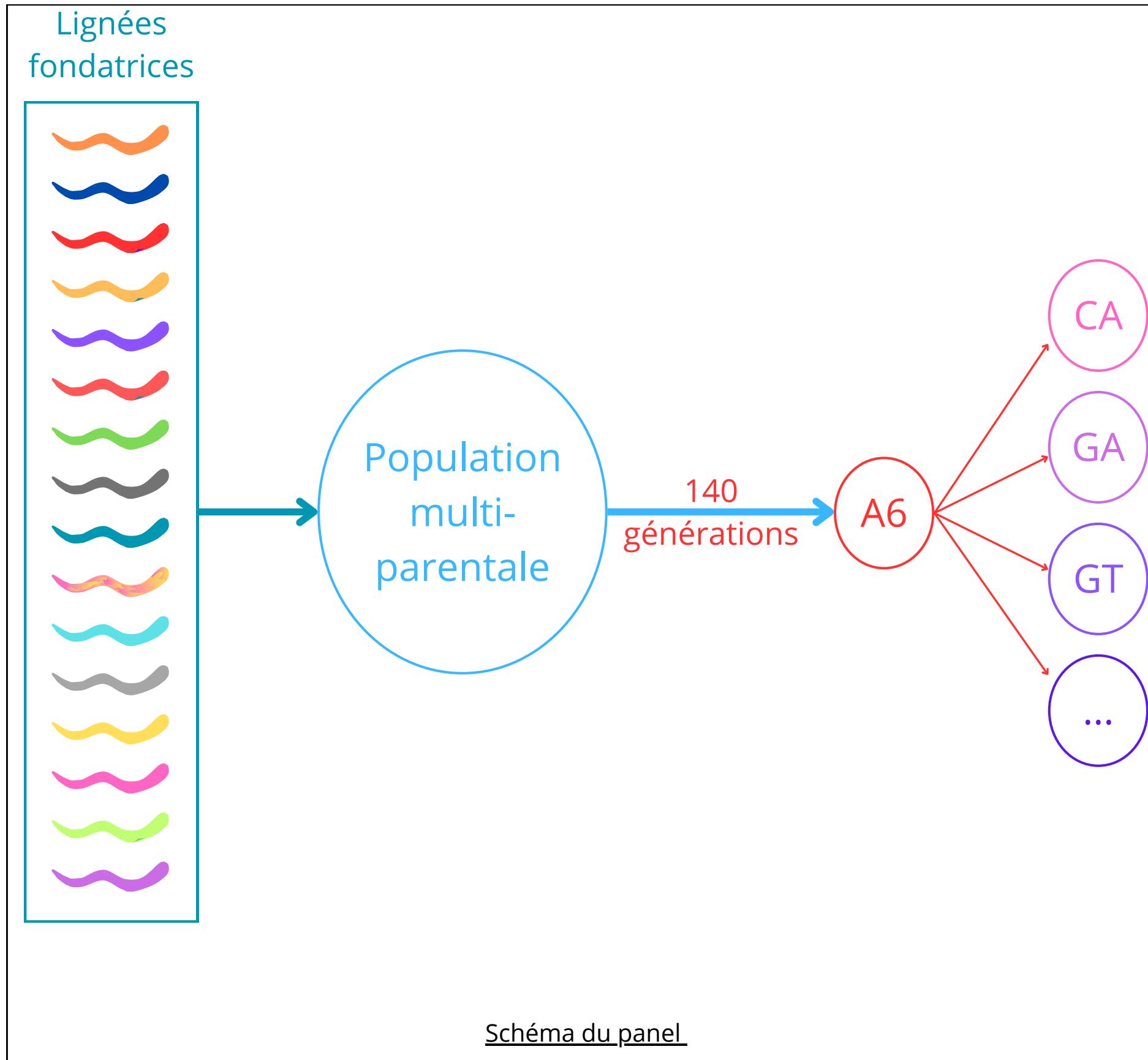
Panel

[Noble, L. & al 2017] [Noble, L. & al. 2021]



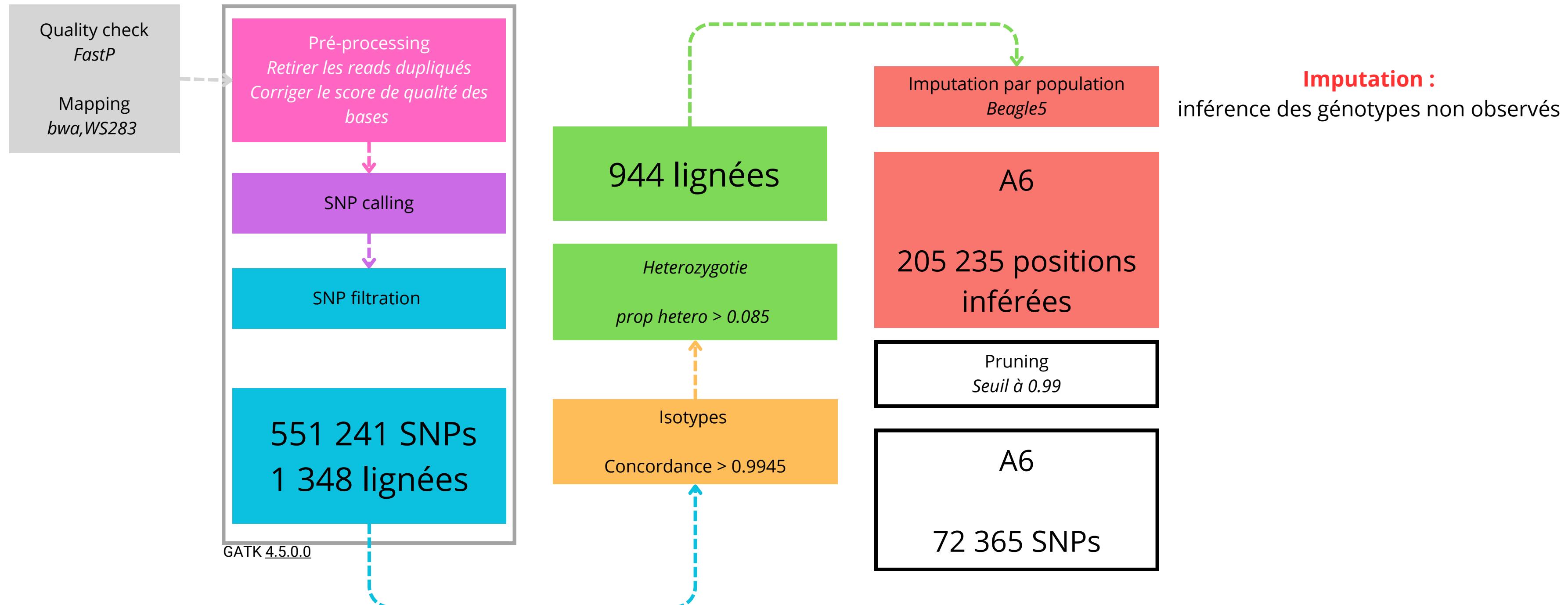
Panel

[Noble, L. & al 2017] [Noble, L. & al. 2021]



- Lignées Recombinantes (RILs)
 - 10 générations minimum d'auto-fécondation
 - Homozygotes
- 1350 lignées séquencées

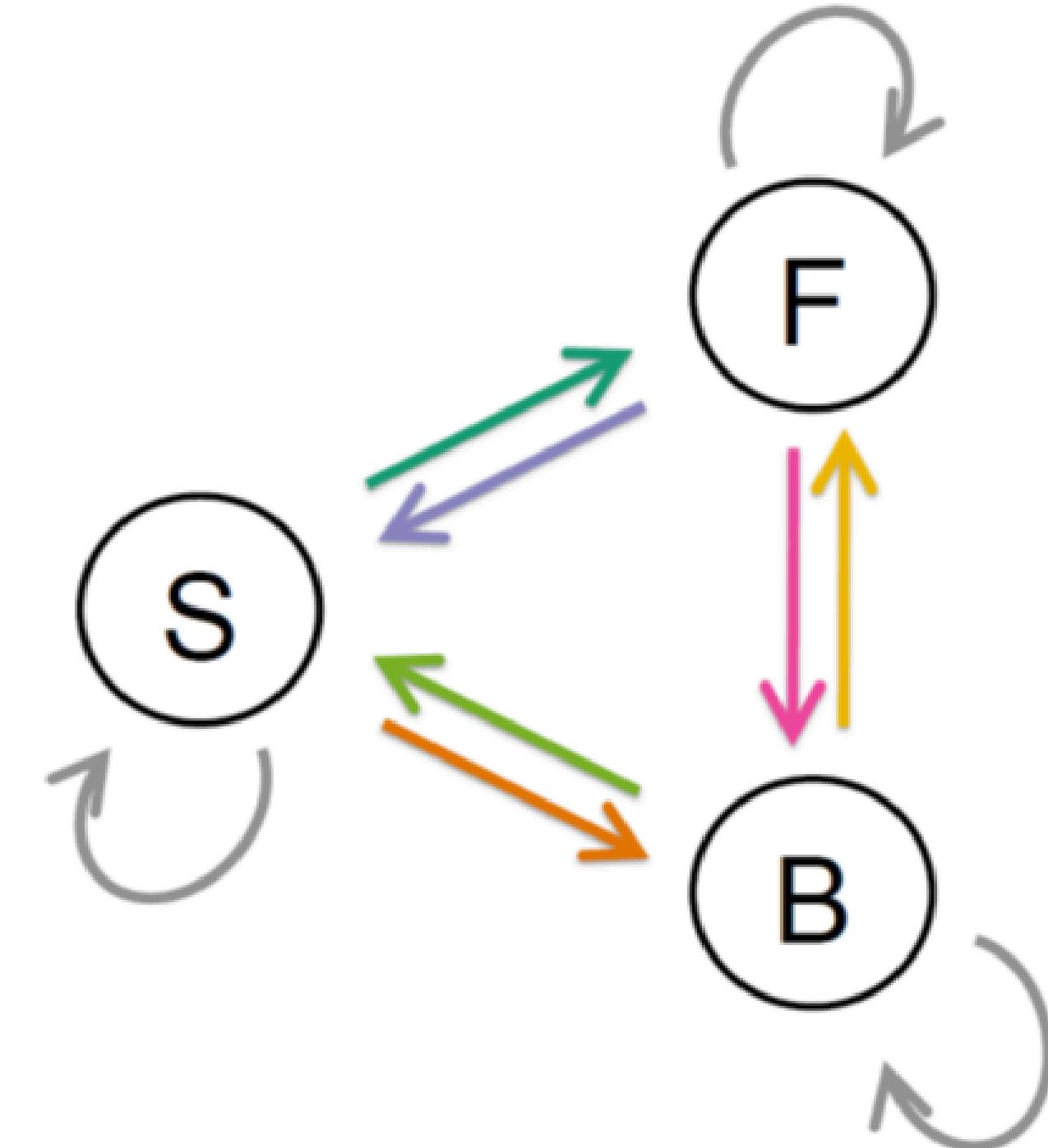
Genotypage



La locomotion

[Mallard, F. & al., 2023]

- Etudié en neurobiologie
- 1D : Chaîne de Markov
 - 3 états
 - 9 taux de transitions
 - 6 indépendants
- Films
 - Au moins deux replicats
- 747 lignées phenotypées
 - NGM/NaCl
 - A6
 - 178 NGM
 - 177 NaCl



Etats et transitions schématisant la locomotion

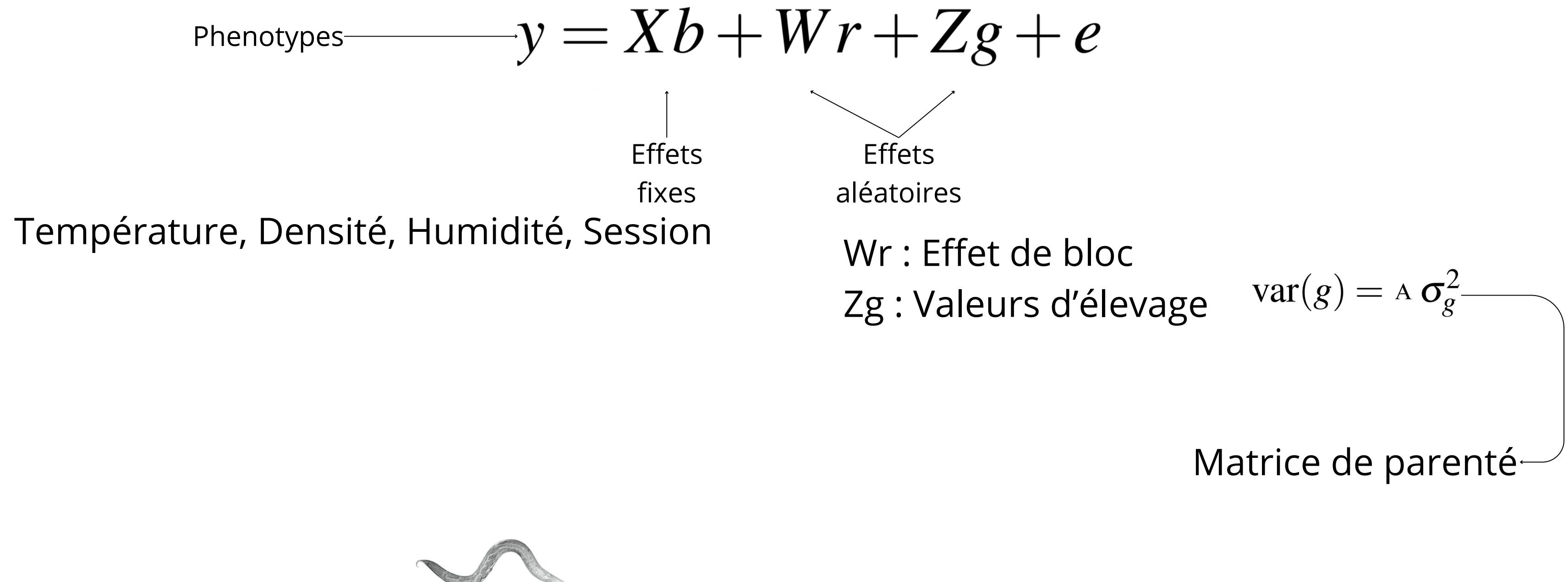


Estimer les valeurs d'élevage (breeding values)

GBLUP

[Legarra, A. 2018]

- MCMCglmm



Prendre en compte la structure de la population

Matrice de parenté

[VanRaden, P.M. 2008]

SNPs

- Equilibre de Hardy Weinberg
- Equilibre de liaison

M individus x SNPs

0, 2

$$Z[i,] = M[i,] - 2(p_i - 0.5)$$

$$A = \frac{ZZ'}{2 \sum p_i(1 - p_i)}$$



Estimer les effets des SNPs

Des VE aux effets des SNPs [Legarra, A. & al 2018]

$$g = \mathbf{Z}\mathbf{u}$$

effets des SNPs

$$\mathbf{Var}(g) = \mathbf{Var}(\mathbf{Z}\mathbf{u})$$

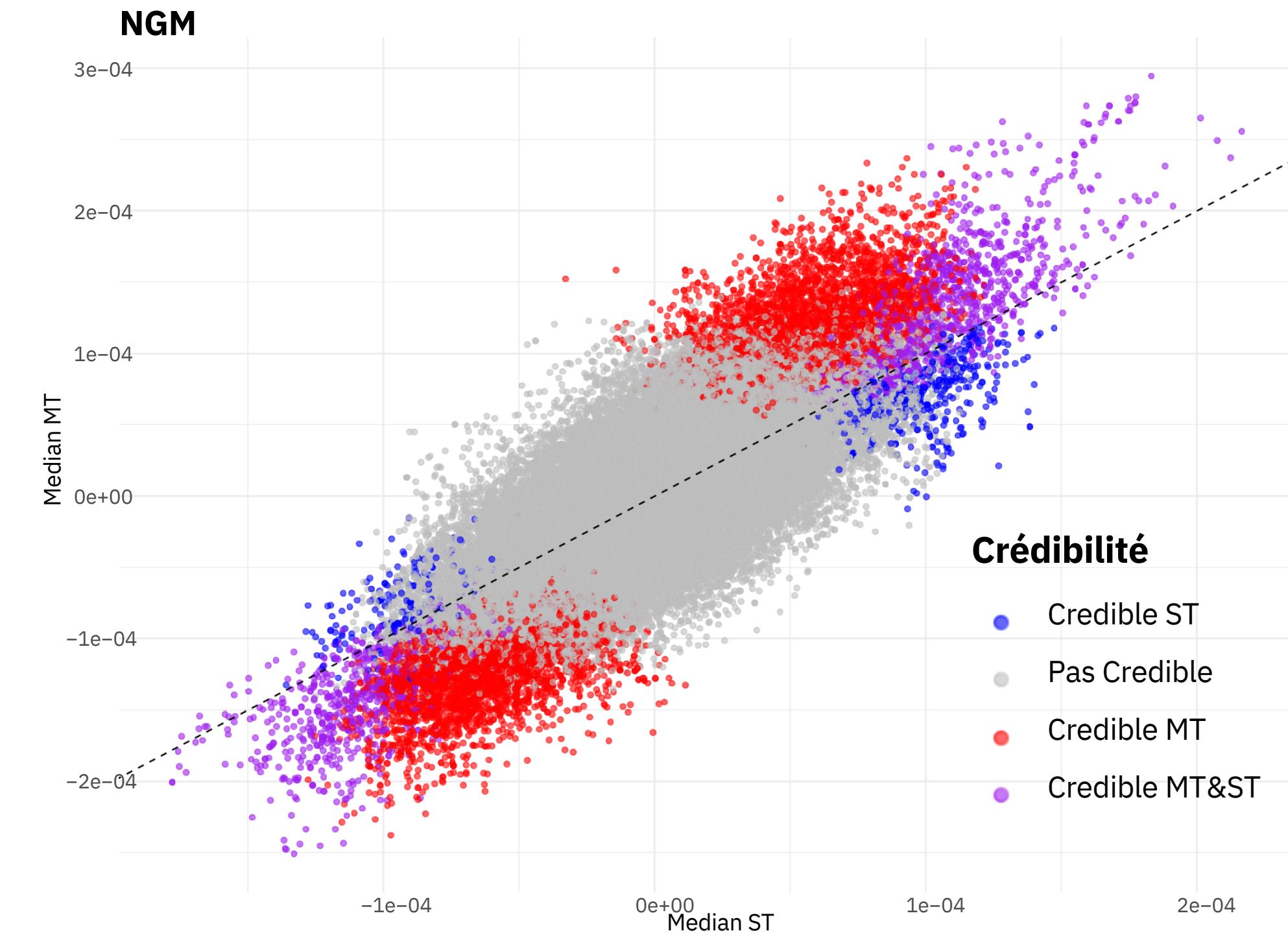
$$\mathbf{u} = \mathbf{M}'(\mathbf{M}\mathbf{M}')^{-1}g$$

- Transformation linéaire des valeurs d'élevage
- Multi-trait : comment se comportent les covariances ?
 - Matrice G



Corrélation des effets des SNPs

- **Crédibles** : IC ne chevauche pas 0
- Corrélation entre les analyses
 - Tendance à être plus élevée pour multivariée

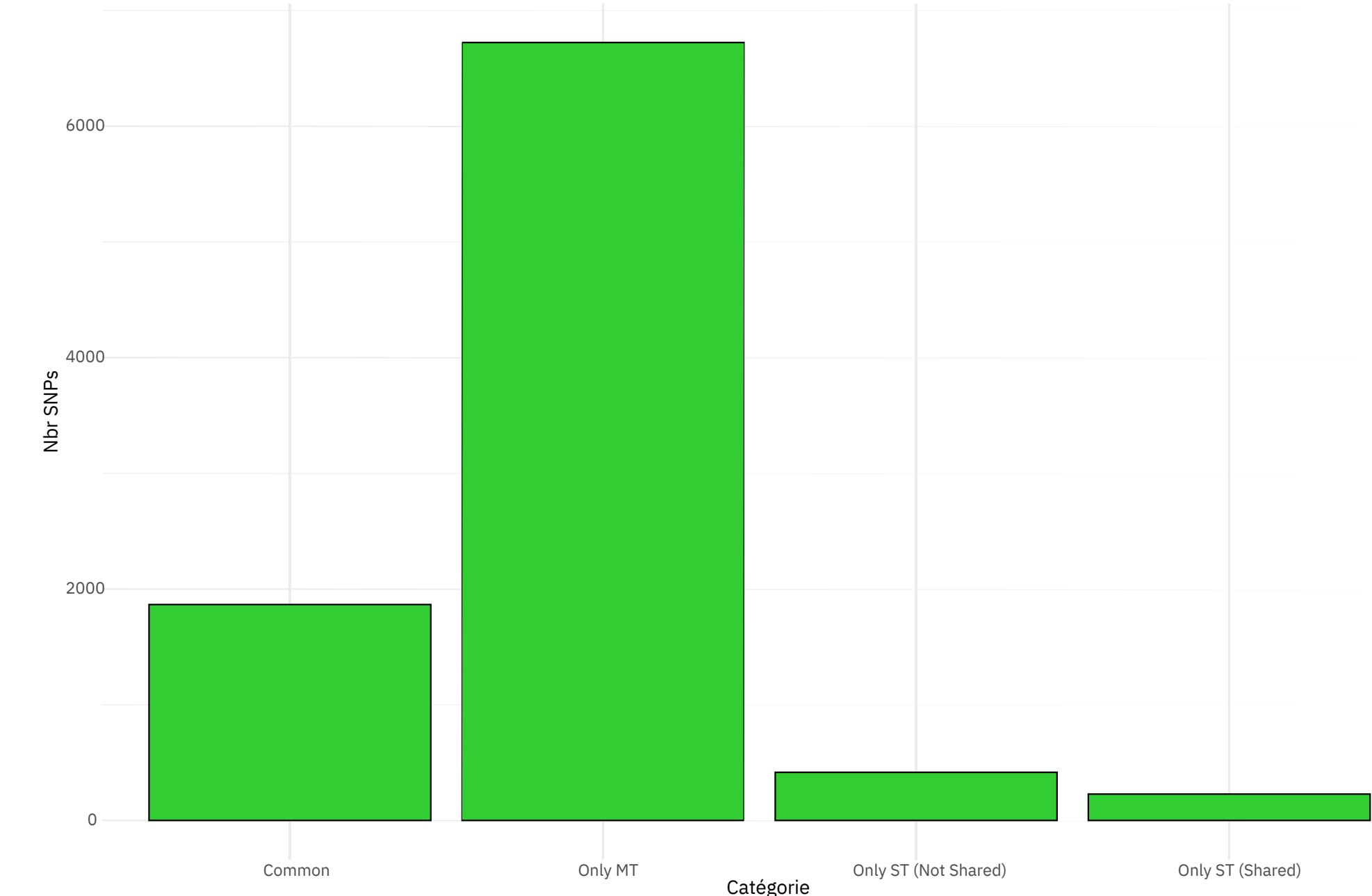
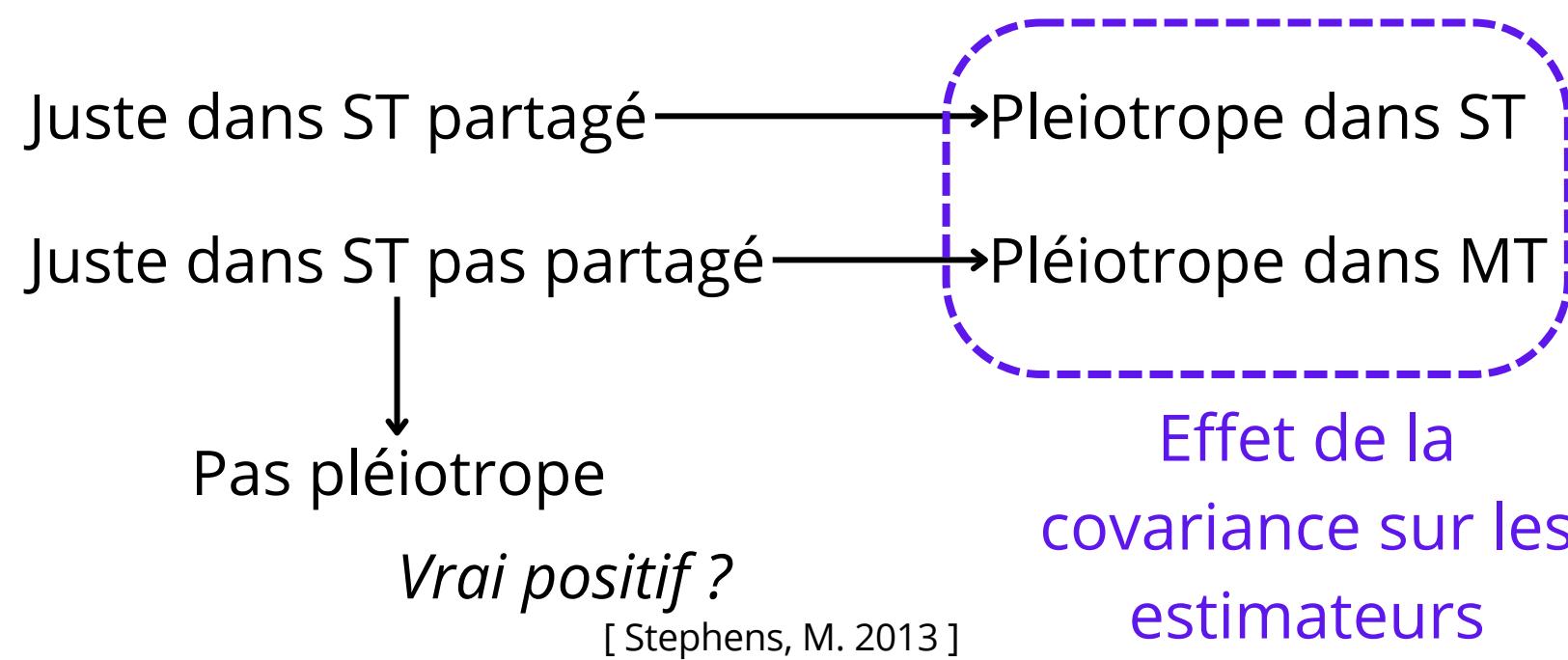


Médiane de la distribution des effets des SNPs à posteroio dans l'analyse
Multi-Trait en fonction de leur médiane dans l'analyse Single-Trait

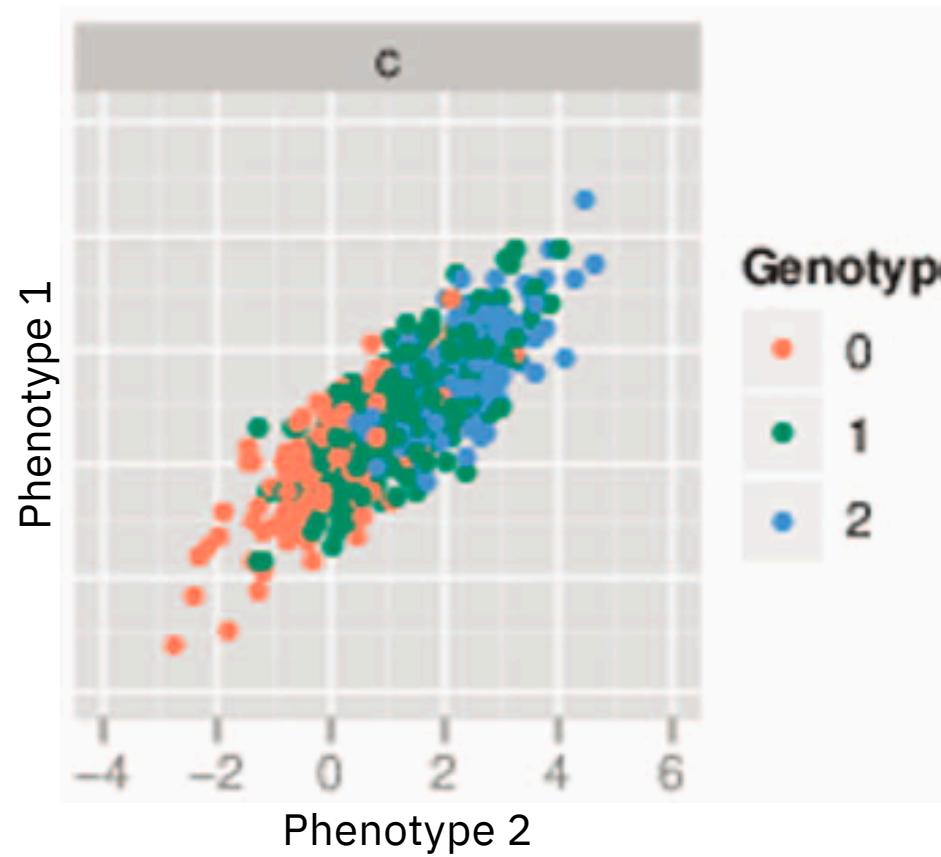


QTLs similaires dans les deux analyses

- +50% des SNPs crédibles retrouvés dans ST dans MT
 - MT et UT semblent cohérentes



Corrélation entre les traits selon le génotype

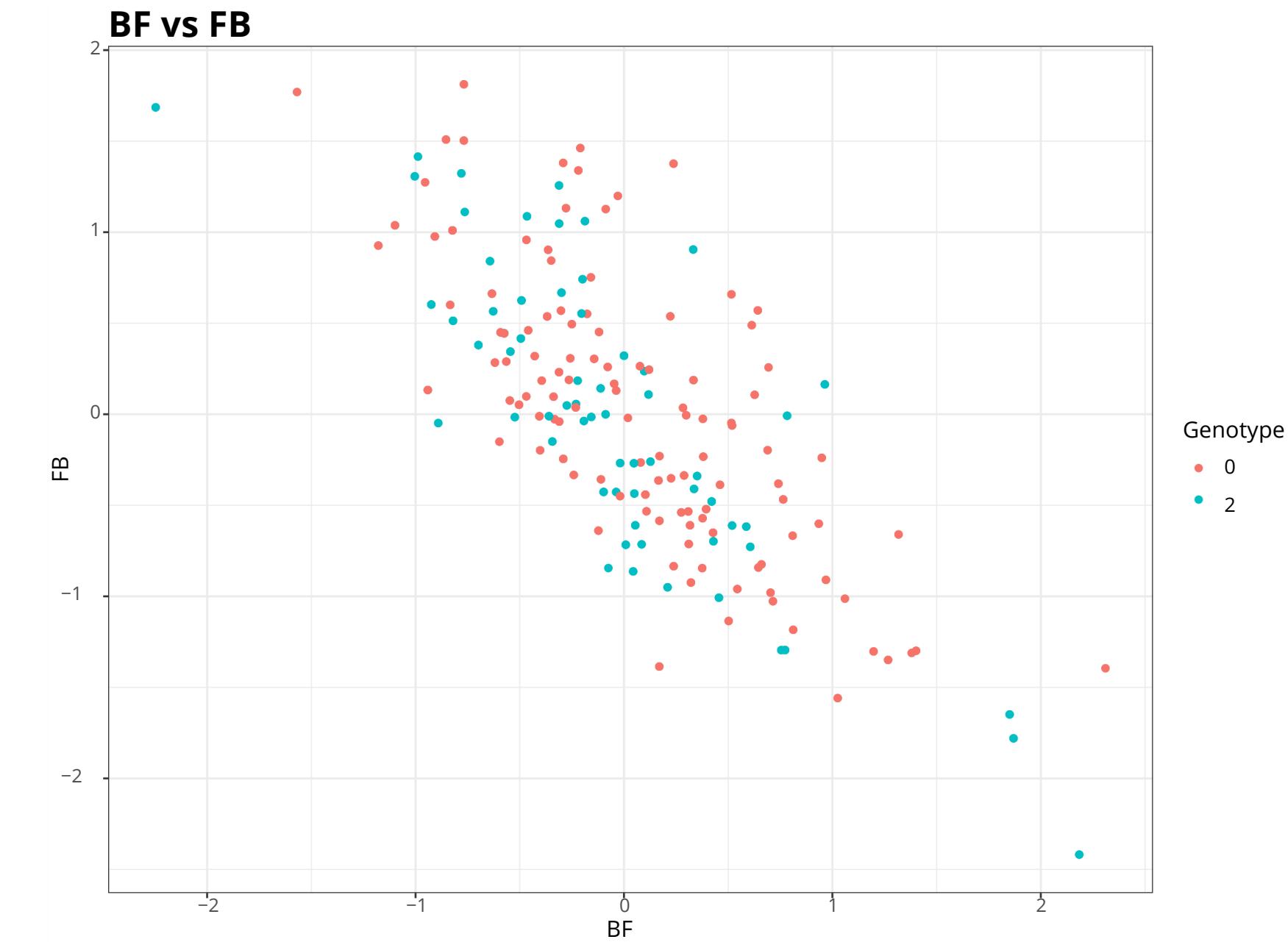


Corrélation entre les phénotypes selon le génotype :
cas dans lequel l'analyse univariée est la plus puissante

[Stephens, M. 2013]

- Pas caractéristique
- Correction Faux Positifs

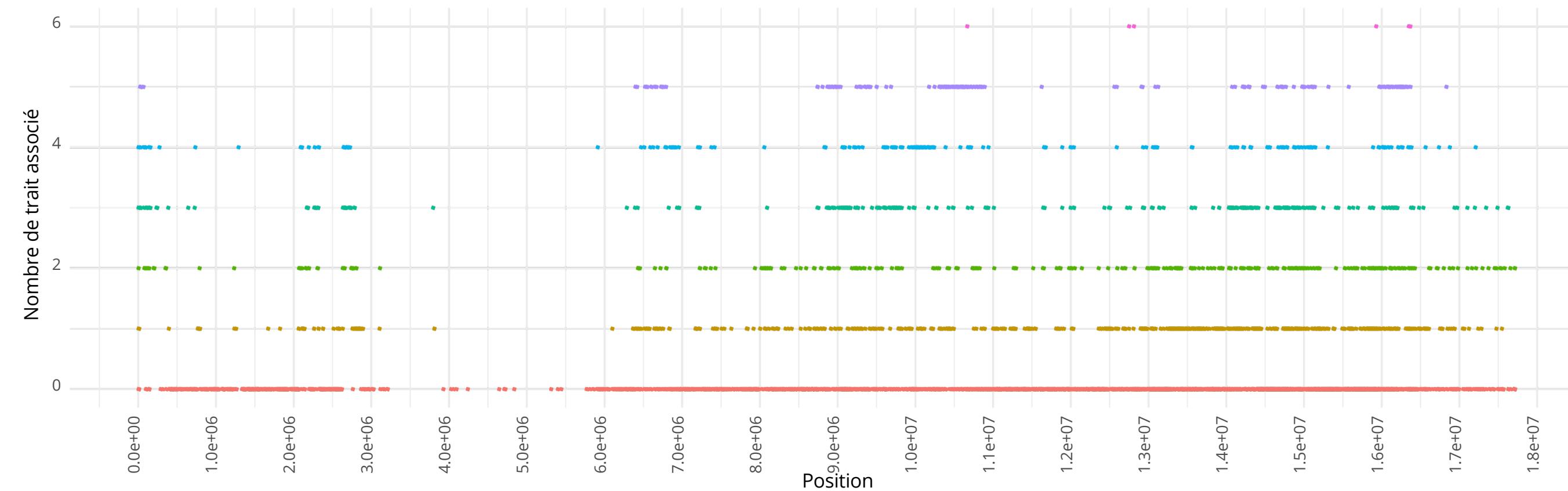
[Stephens, M. 2017]



Corrélation entre les phénotypes FB et BF selon le génotype



Pléiotropie au sein d'un environnement



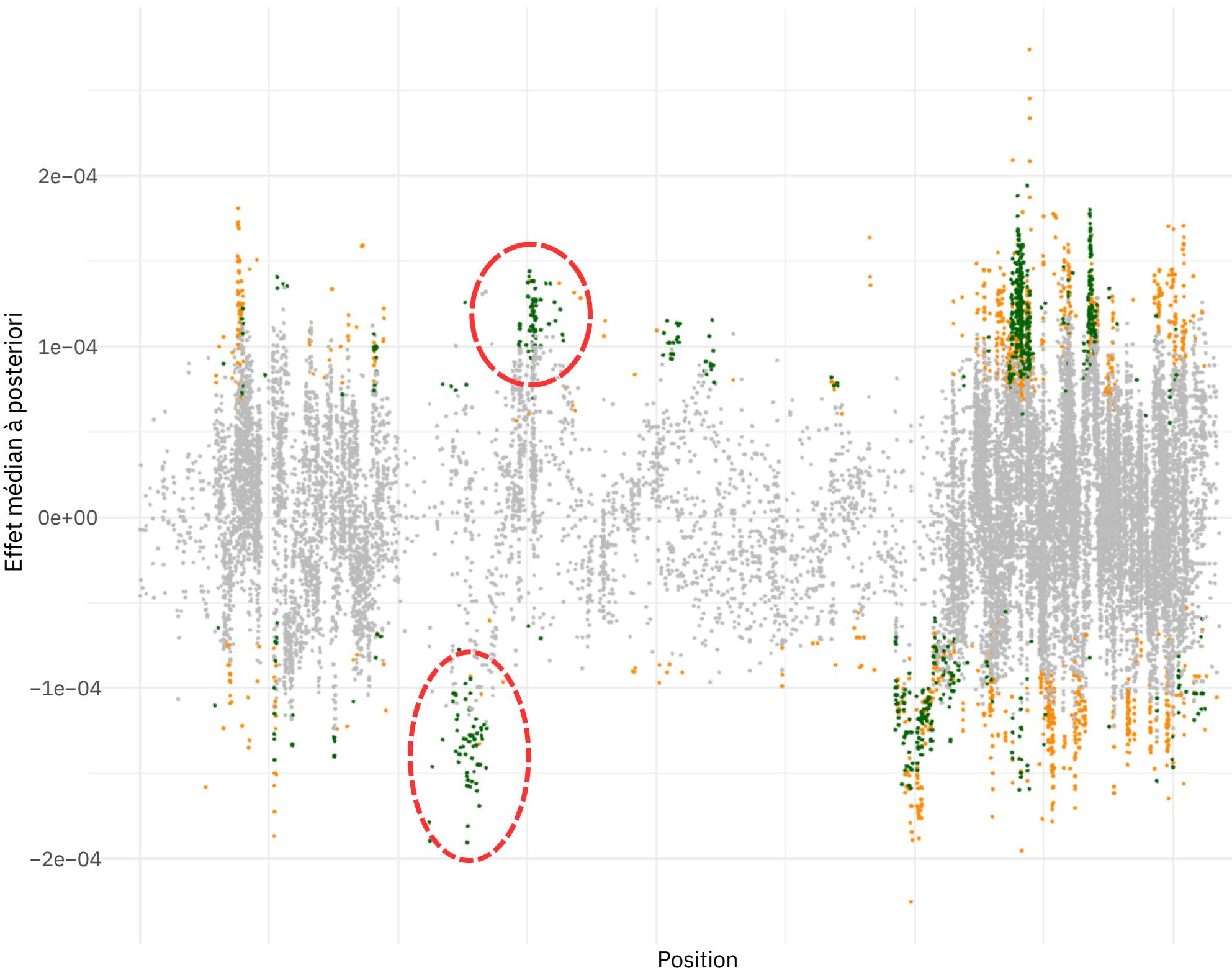
Répartition des SNPs le long des chromosomes selon le nombre de trait auxquels ils sont associées
Chromosome X, condition NGM



Pléiotropie au sein d'un même environnement



Comparaison des SNPs crédibles



- Majoritairement similaires
- Quelques nouvelles régions en NaCl

Variation génétique cryptique :
pas ou peu d'effet en conditions normales, mais qui devient impactante en conditions atypiques.

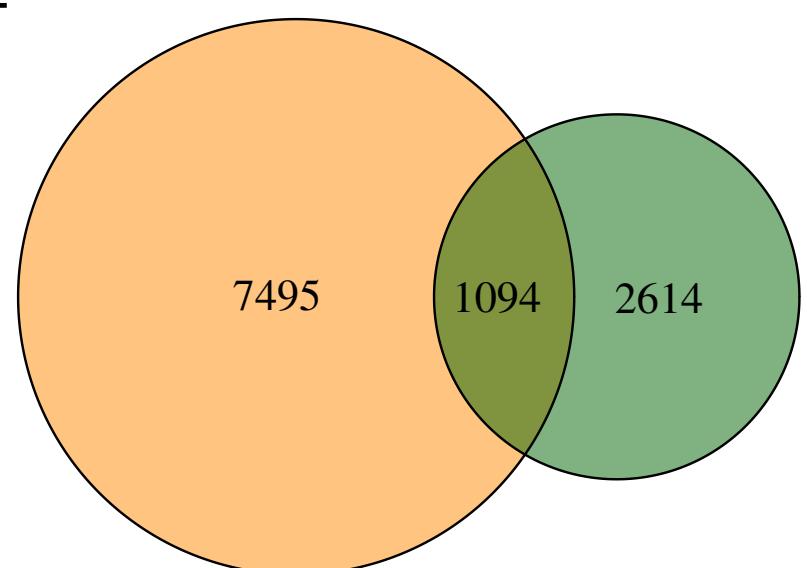
[Paaby, A. & Rockman, M 2014]

"Manhattan" Plot des médianes des effets des SNPs à postériori dans les conditions NGM et NaCl



Pléiotropie entre environnements

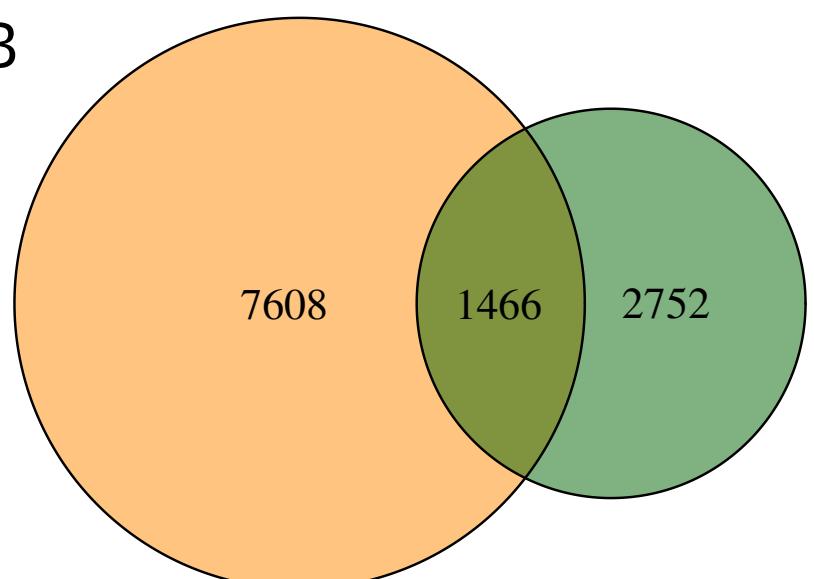
BF



SB



FB



Crédibilité

- Pas Crédible
- Crédible NGM
- Crédible NaCl

Répartition des SNPs crédibles entre les conditions



QTLs différents

- Covariance entre environnement semblent être du aux liaisons



Traits peuvent évoluer indépendamment entre environnement
=> Plasticité



Points clefs

- Matrice G très corrélées avec et sans matrice de parenté
 - Prédiction du phenotype
- Ilots hétérozygocité : région particulière chromosome V
 - *L'échelle des SNPs est elle suffisante ?*
- Analyse multivariée cohérente avec univariées
- Locomotion hautement polygéniques
- Correlation génétique entre environnement semble due aux liaisons



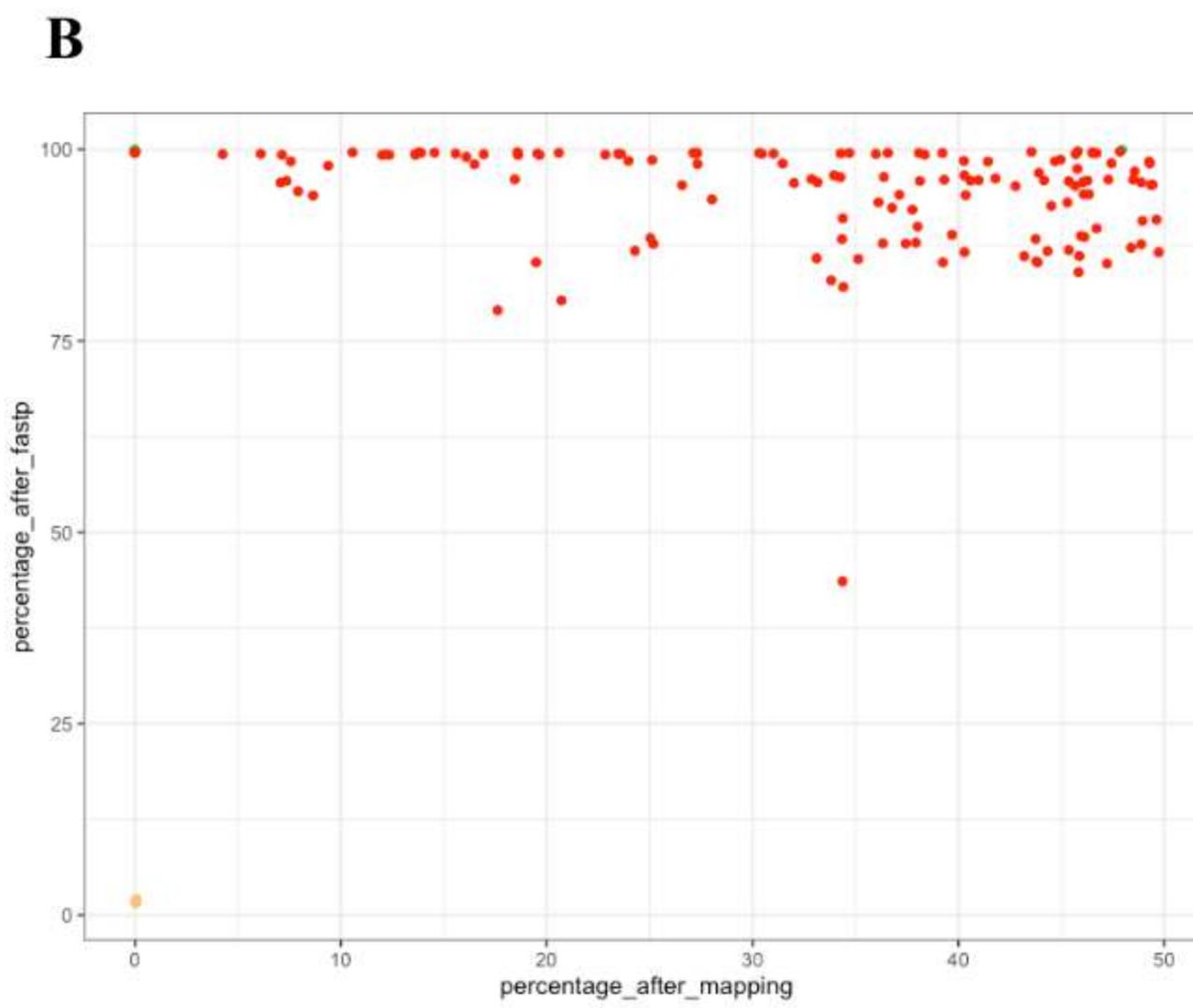
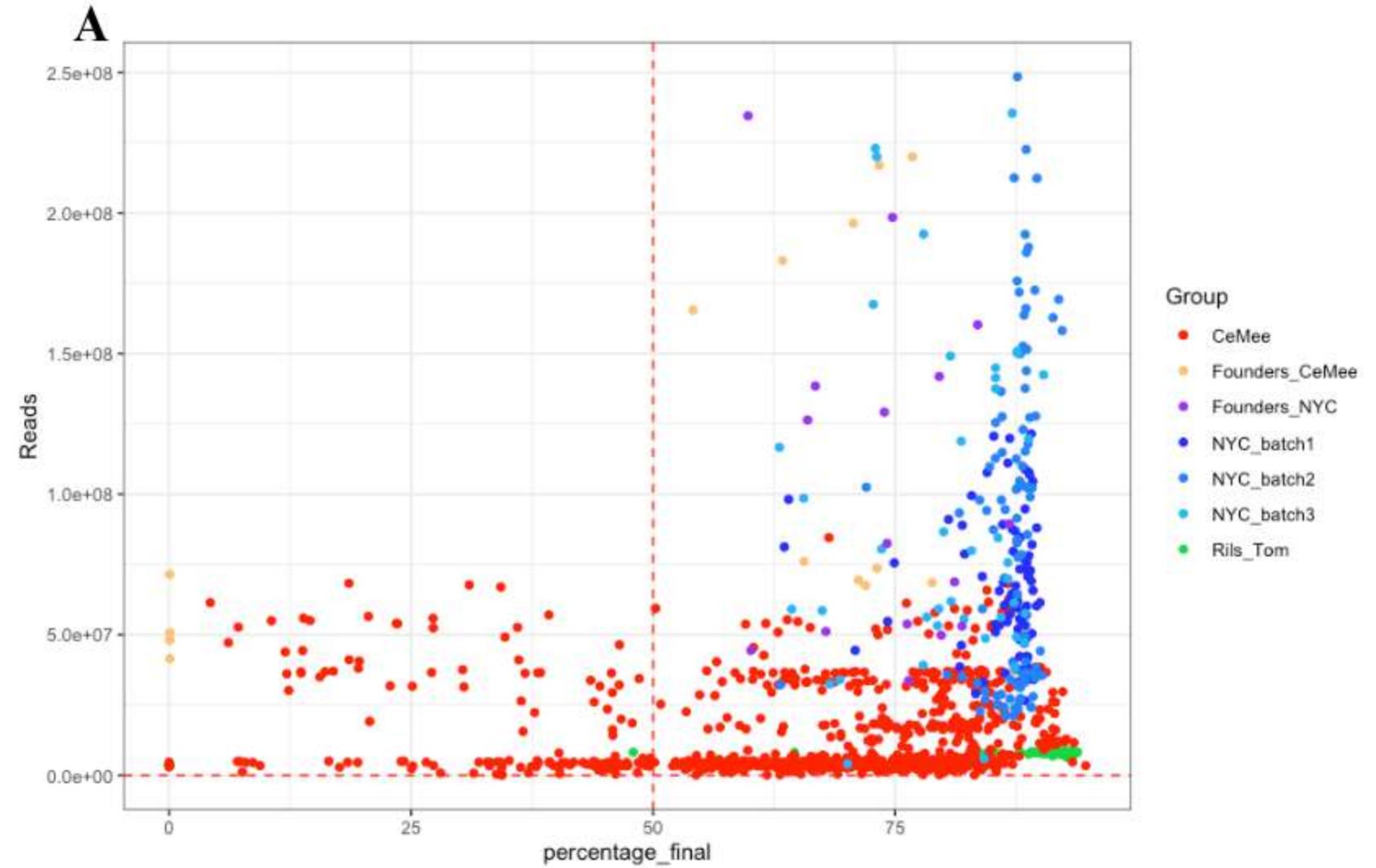
Perspectives

- Comparer aux populations évoluées
 - Tester la causalité recombinaison sur l'évolution de la corrélation génétique entre les traits
- Architecture adaptative
 - Fréquence allélique de départ
 - Prédire les trajectoires de fréquence des allèles dans différents environnements

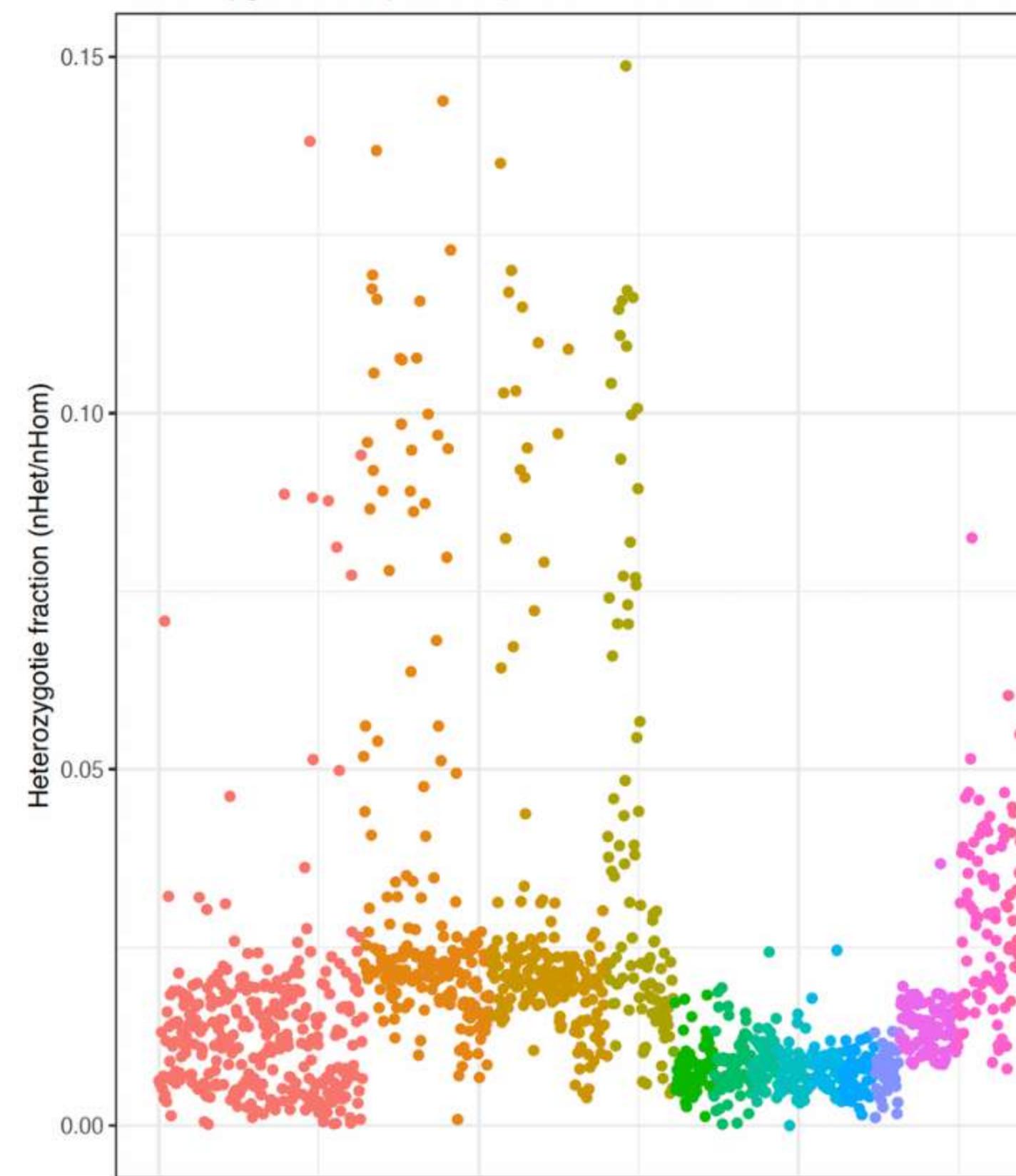


A black and white photograph of a C. elegans nematode, a small roundworm, resting on a textured, light-colored surface. The worm's body is elongated with distinct, rounded segments and a slightly darker, more translucent head region. It is positioned horizontally across the center of the frame, with its body curving slightly downwards towards the bottom left.

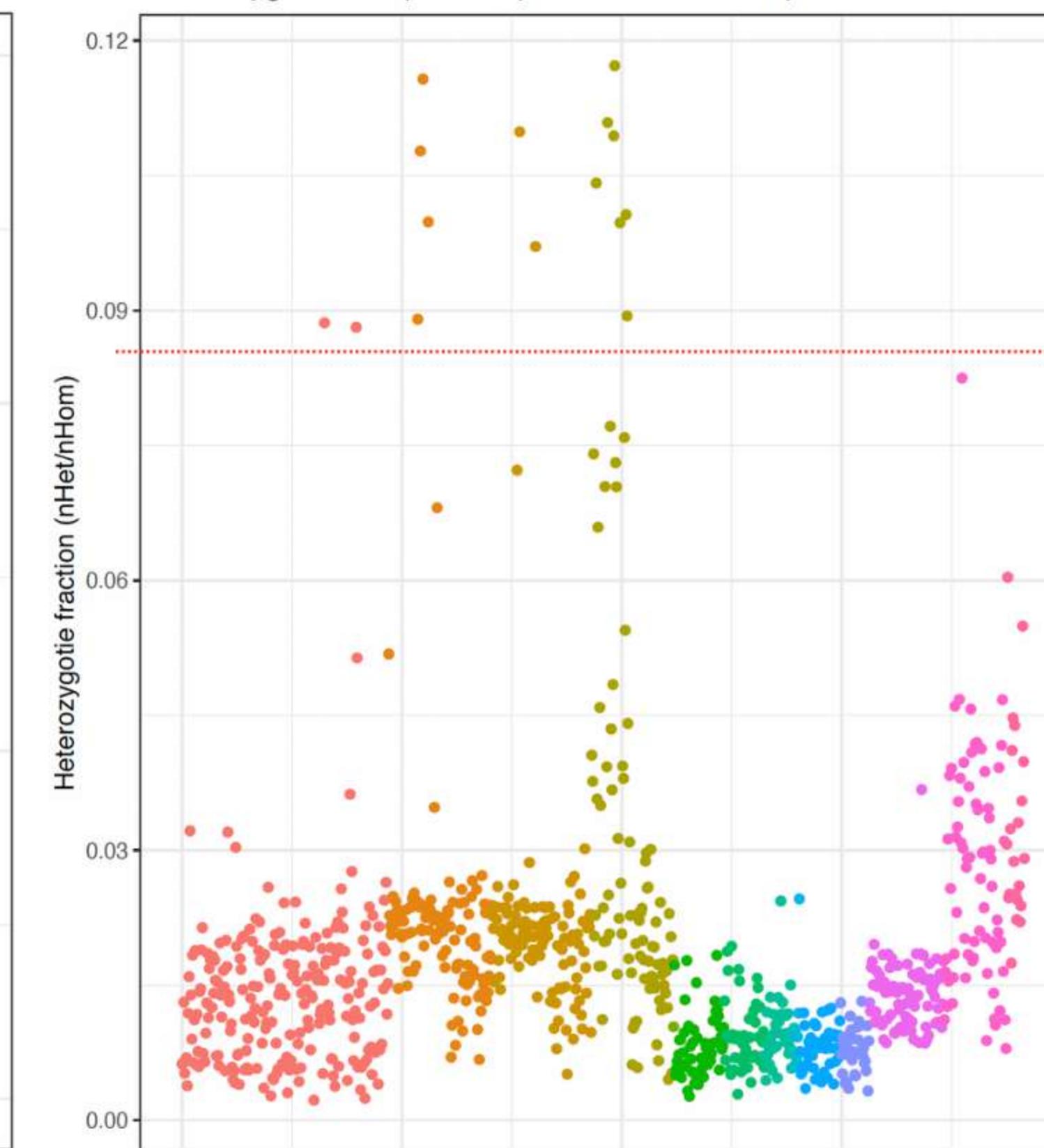
Merci pour votre attention



Heterozygotie rate per sample AllLines.bF



Heterozygotie rate per sample AllLines.bF.sampleRemoved



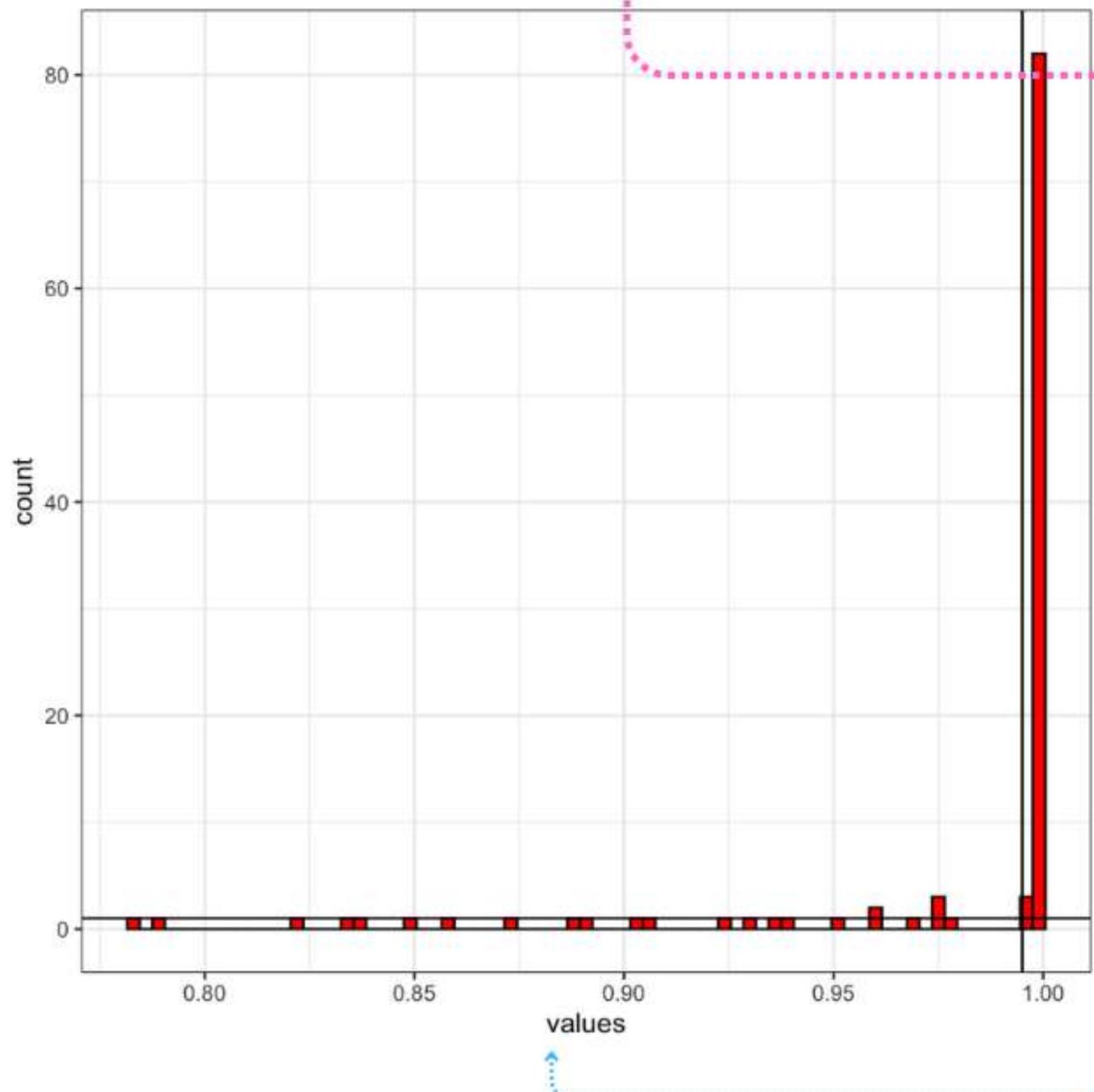
population

- A6
- CA1
- CA2
- CA3
- EEV
- GA1
- GA2
- GA4
- GM1
- GM3
- GT1
- GT2
- LR1
- LR3
- SMR2
- SMR4

All lines

Calling SNPs

Similar lines : Isotype groups

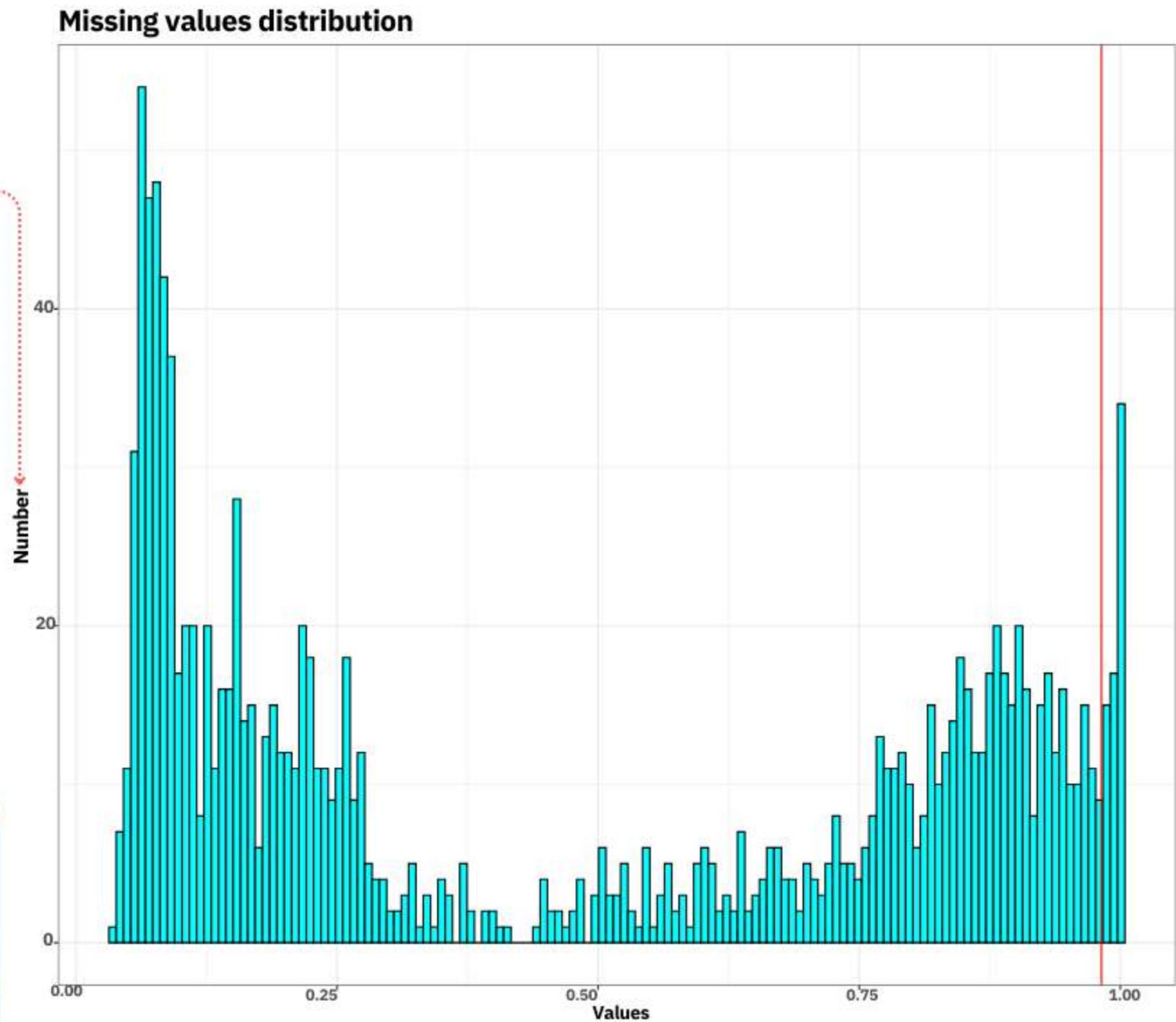


Removed lines with too many missing values

- Cutoffs set at 0.982
- Removed 68

Set cutoffs from known isotopes (ie the 121 pairs sequenced twice)

- Set a cutoffs as two lines are similar if the have concordance > 0,9945
- Removed known isotopes with lower values (48)

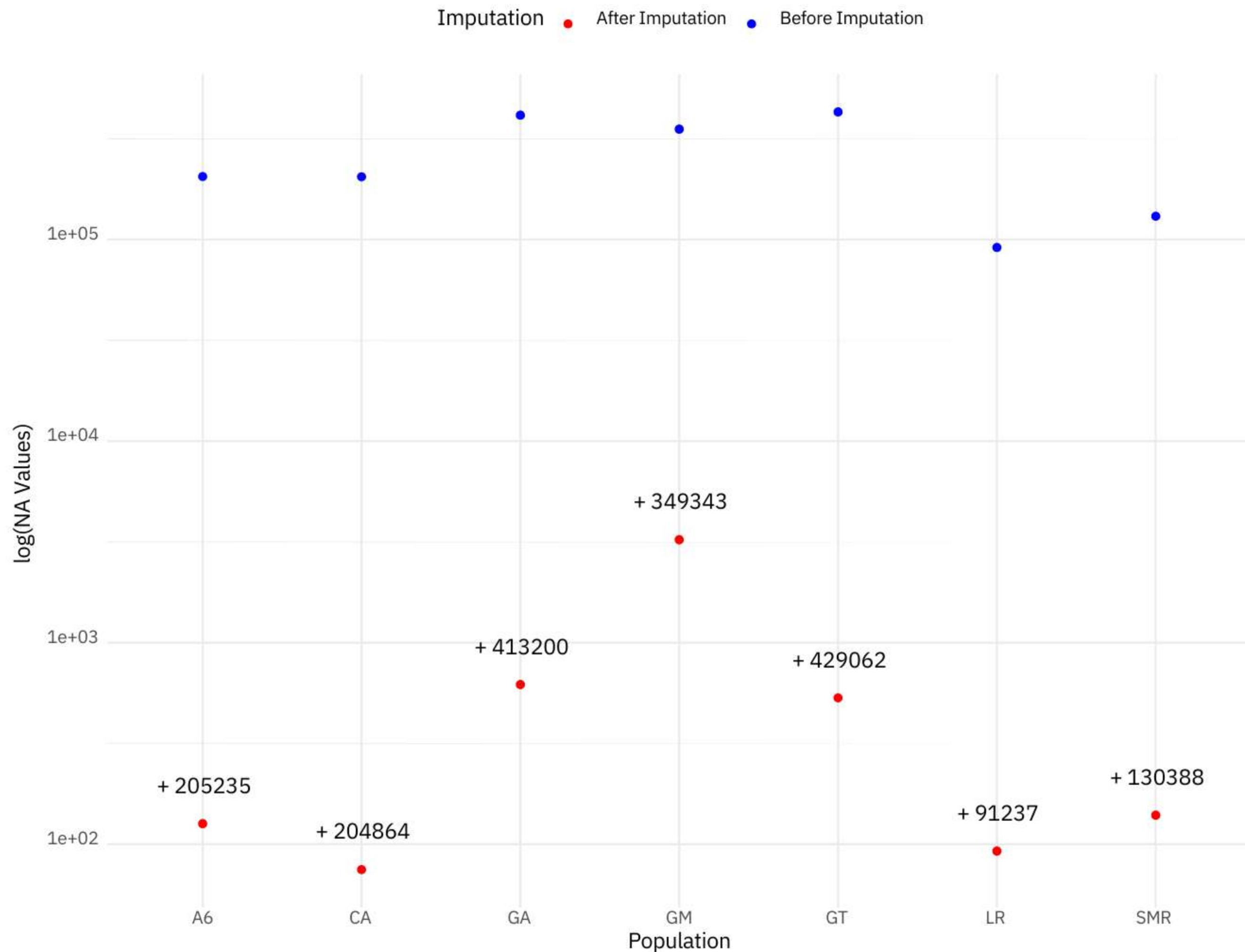


Find isotype groups

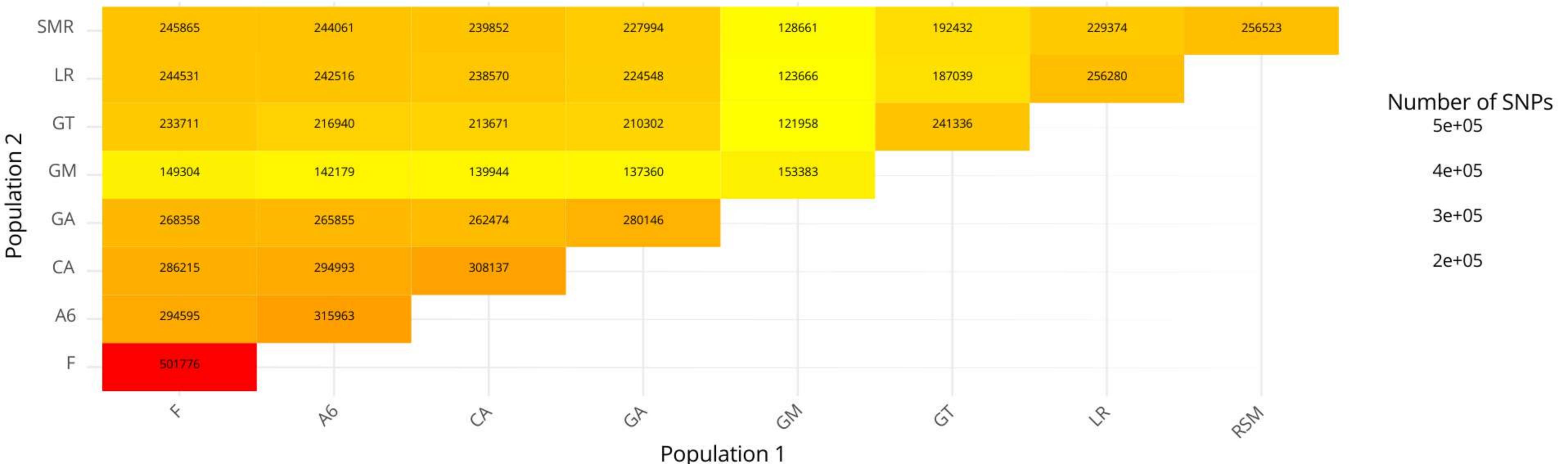
- Main difficulty : reciprocity, find true friend...
- Some lines shared similarity between several groups => removed them (17)

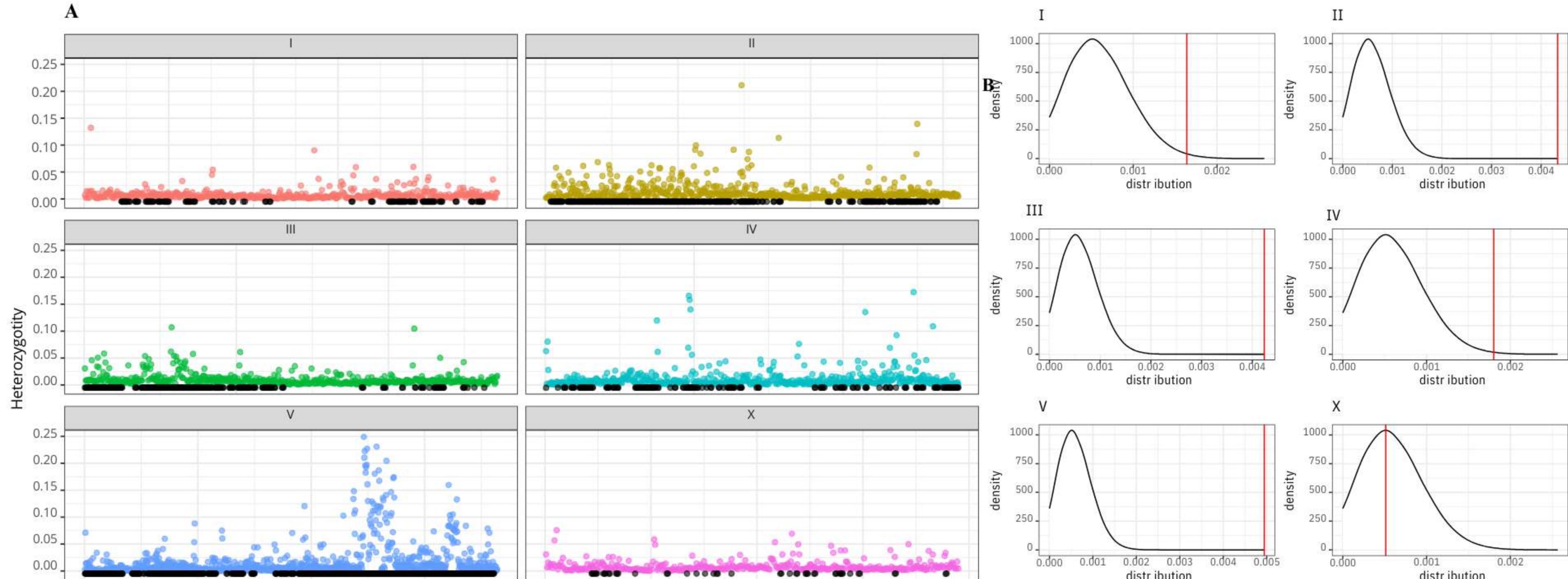
390 lines implied in 135 isotype groups : 255 lines removed
Left : 959 lines

Comparison of NA Values Before and After Imputation

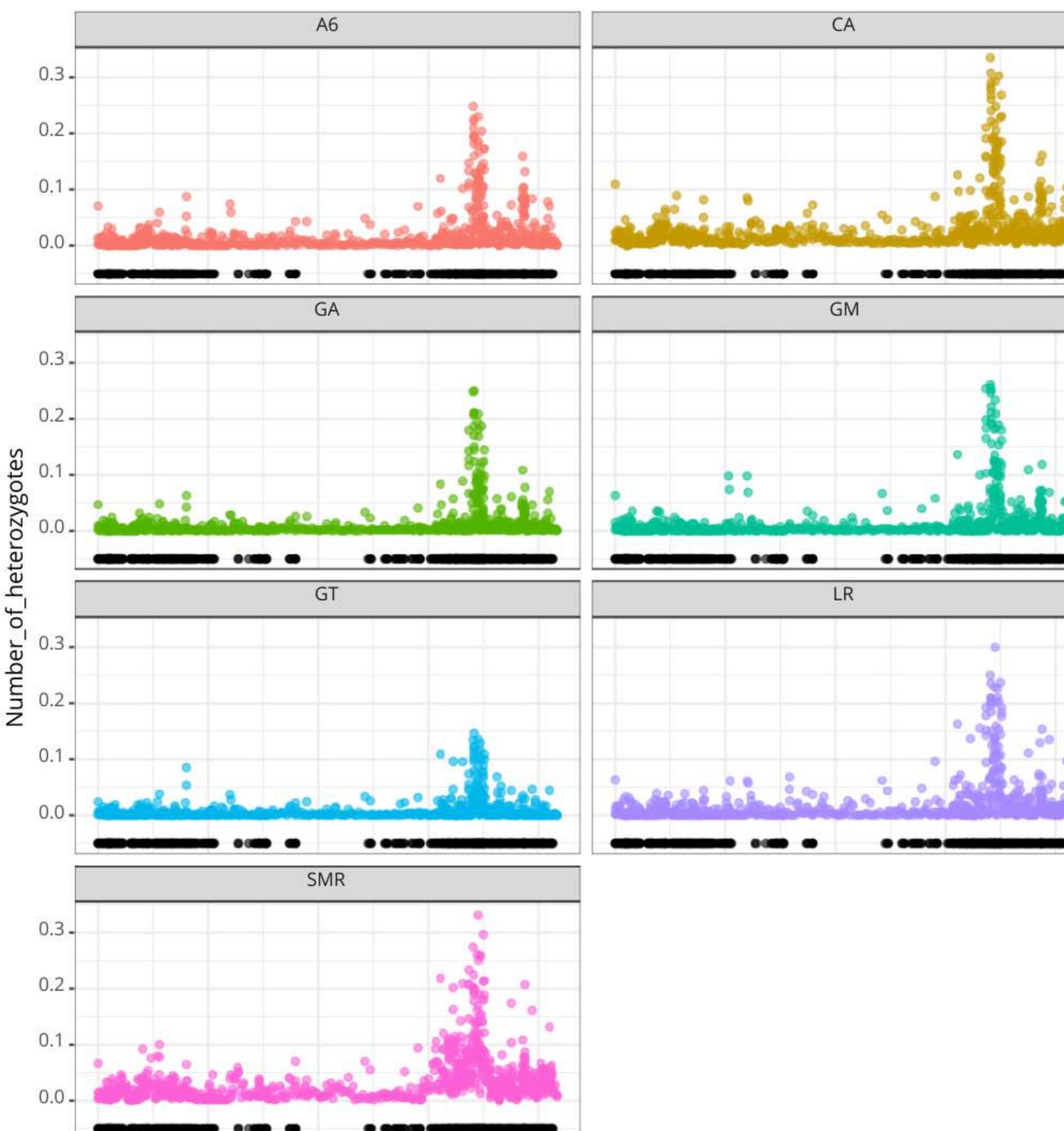


Heatmap of Shared SNPs Between Populations

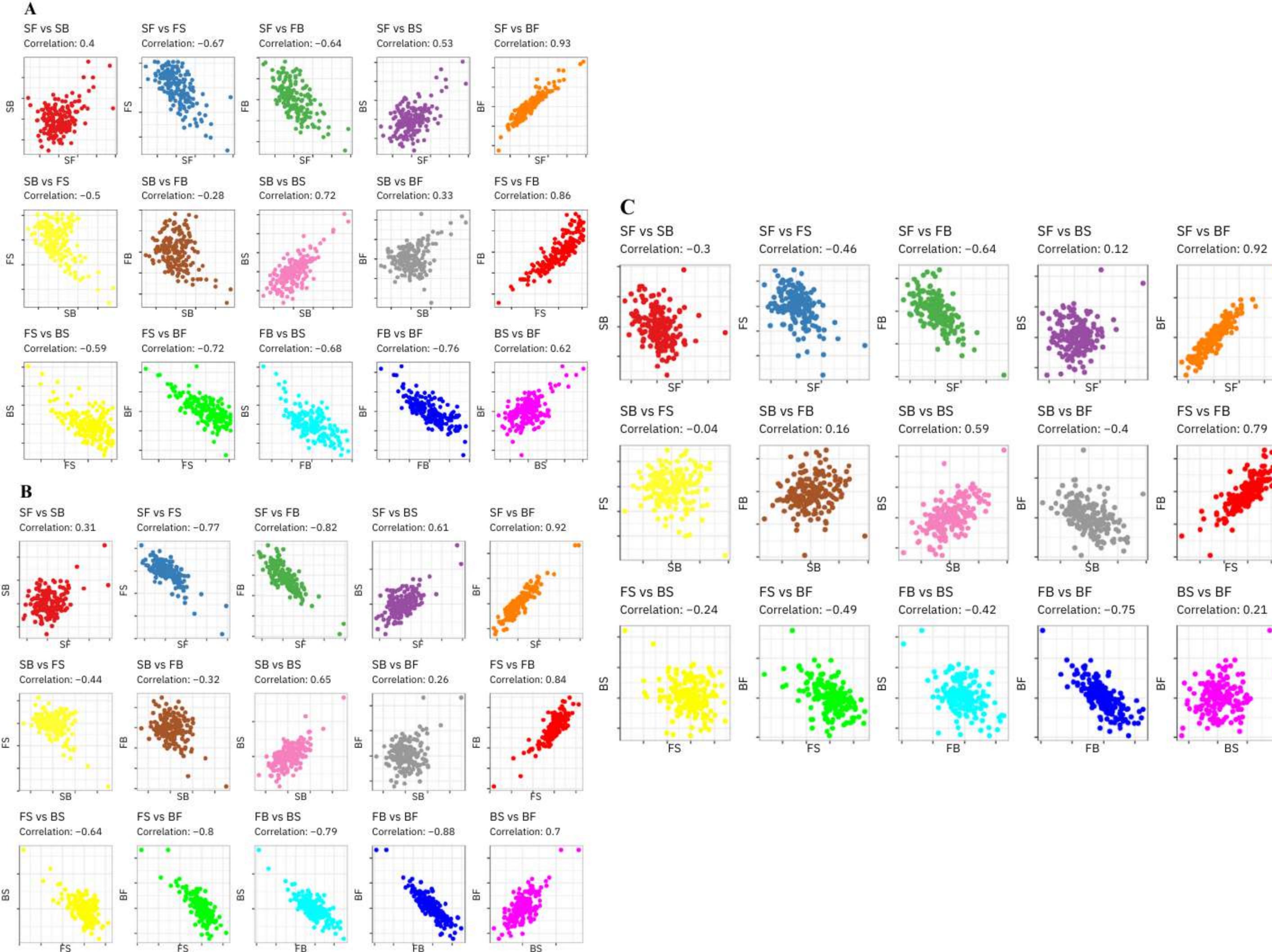


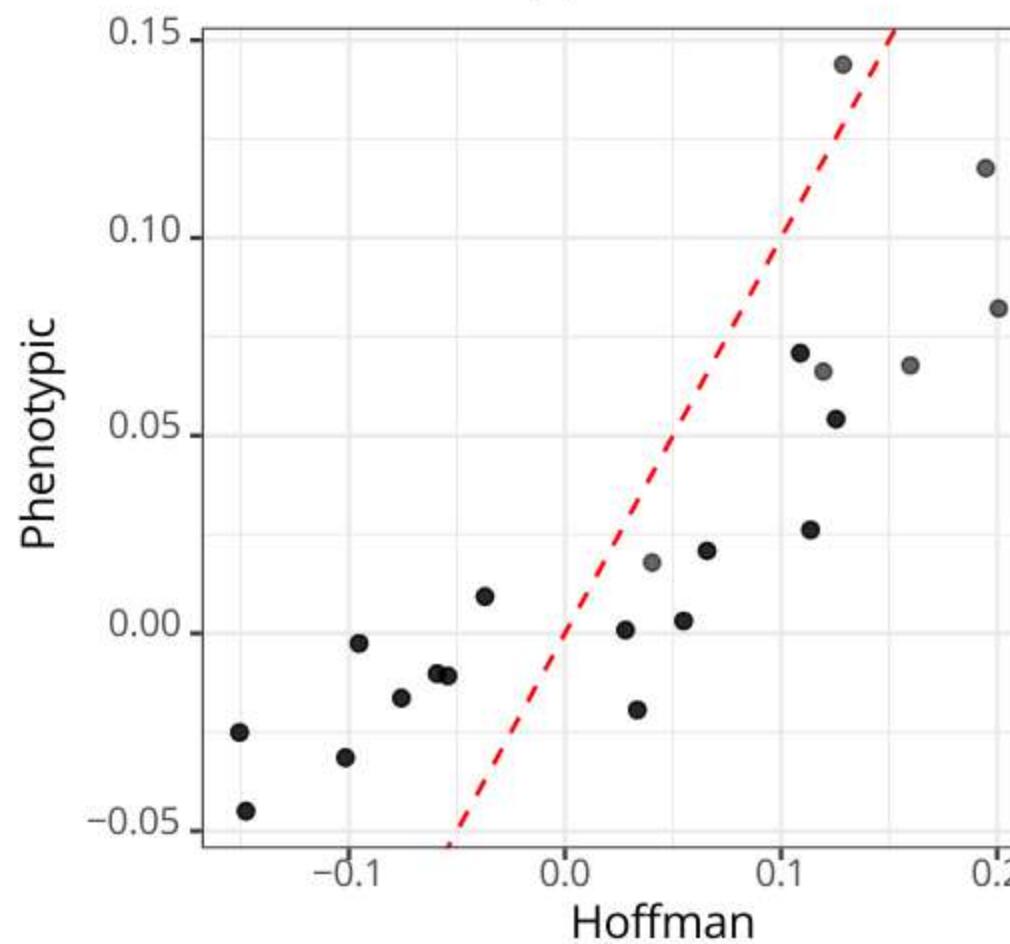


- Fenêtre de 500 SNPs
- **Régions hyper divergentes** : segment du génome avec un taux de variation bien supérieur à la moyenne

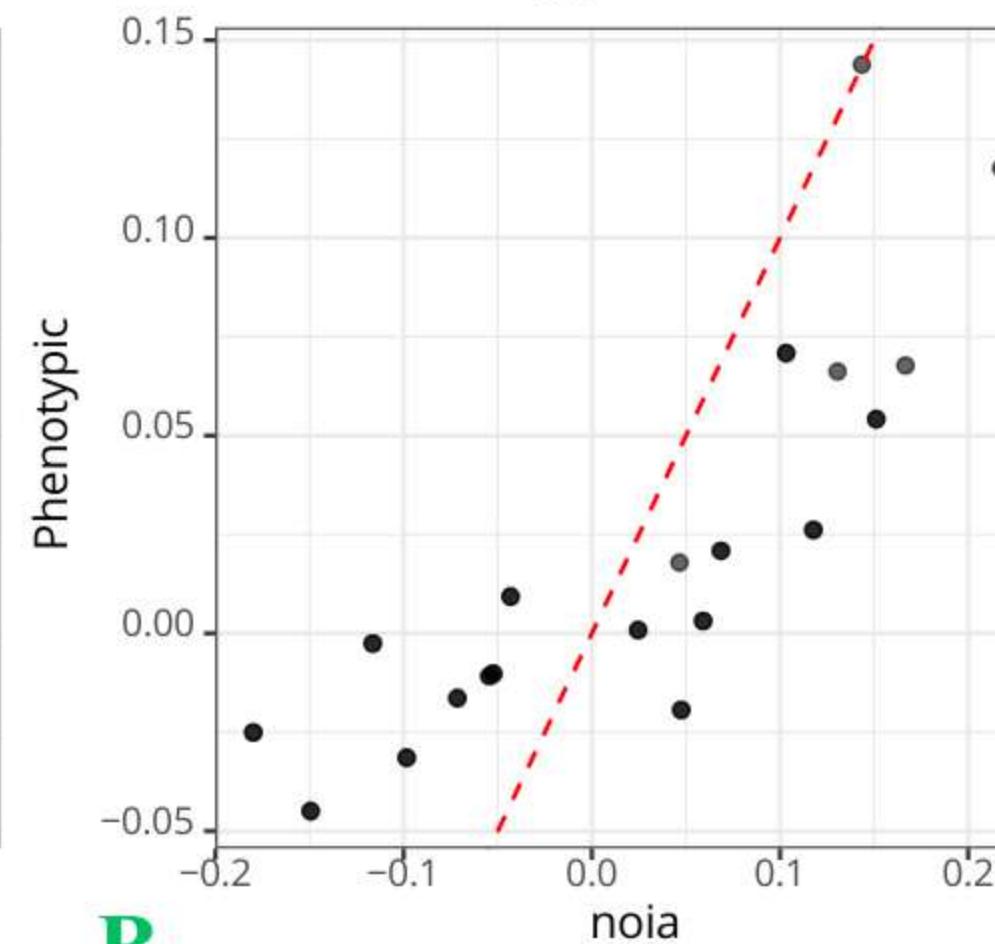


- - Homozygotes peut être lethal
Peu probable, sélection devrait le retirer de la population
- Variants structuraux
deletions, duplications, nombre de copie de gène, insertions, inversion translocation,
- Problème d'alignement sur le génome de référence
- Séquençage avec des longues lectures, pangénom...

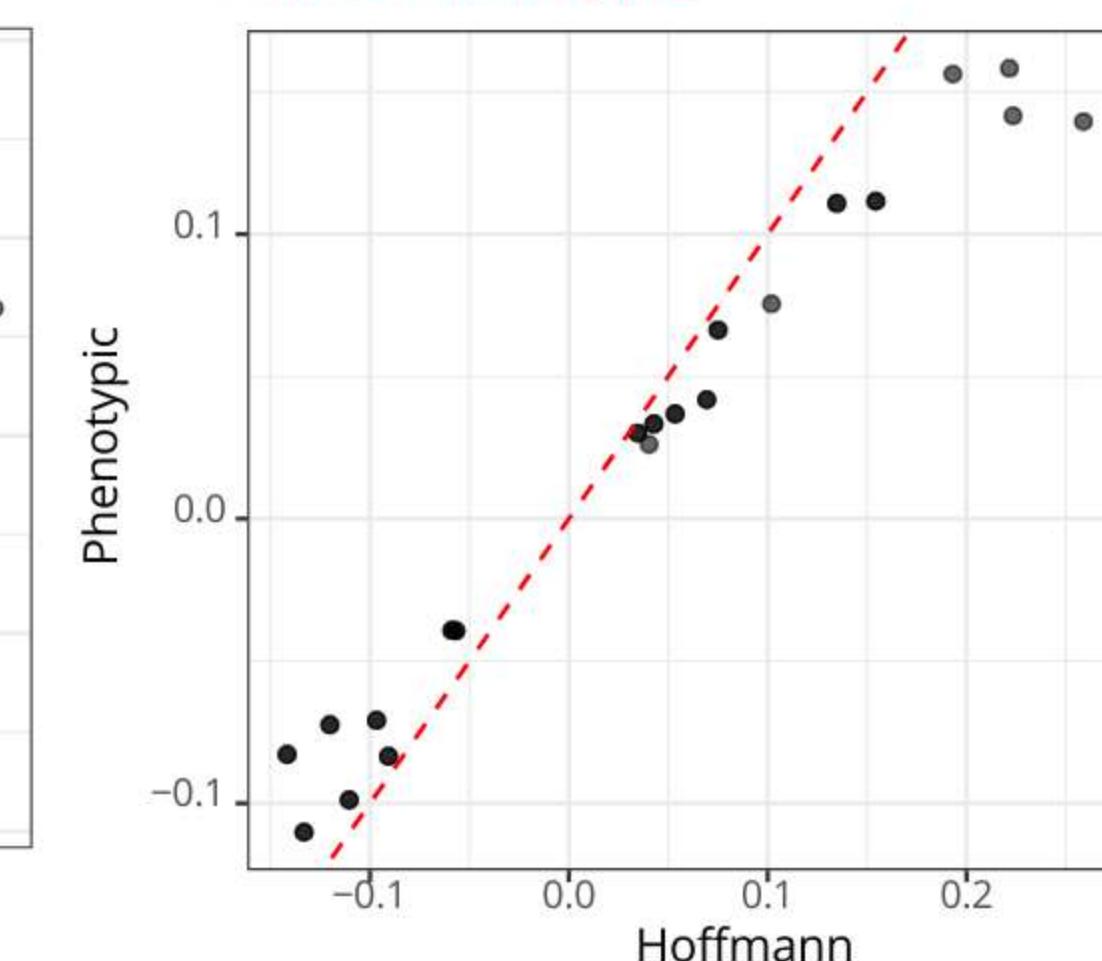


A A6 VS Phenotypic

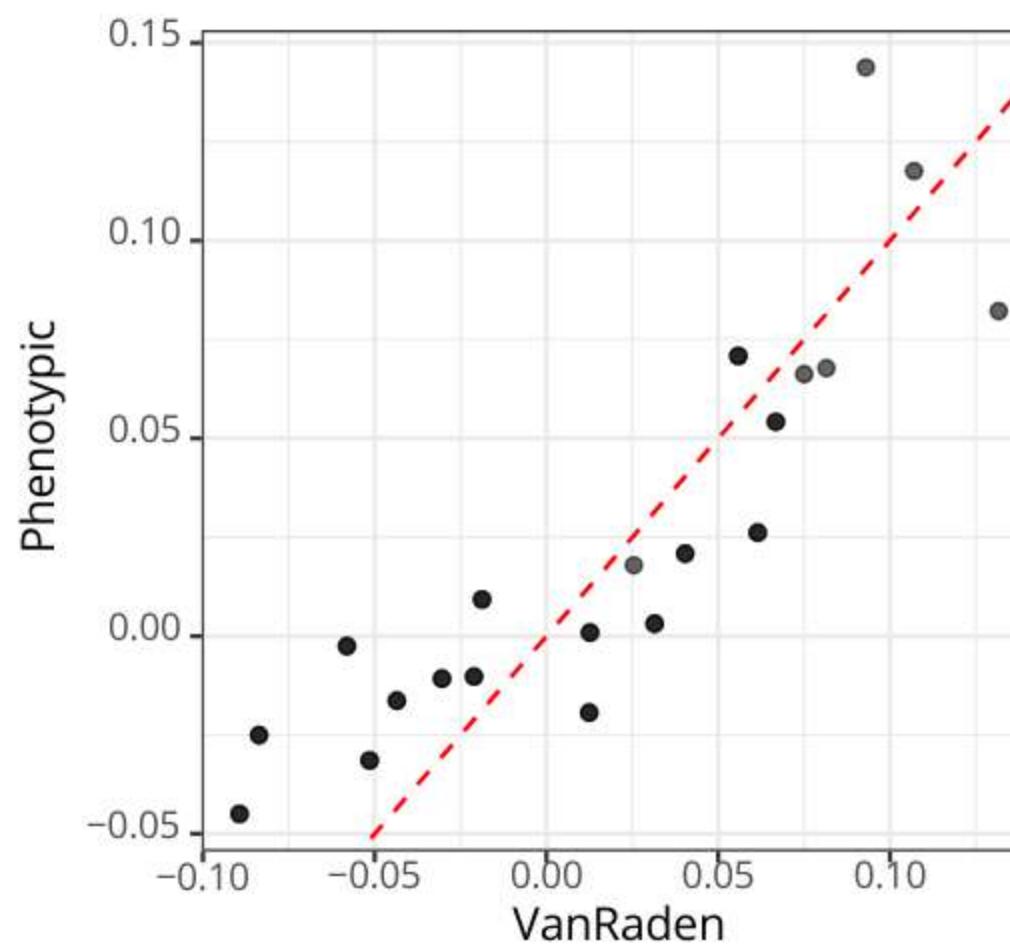
A6 VS Phenotypic



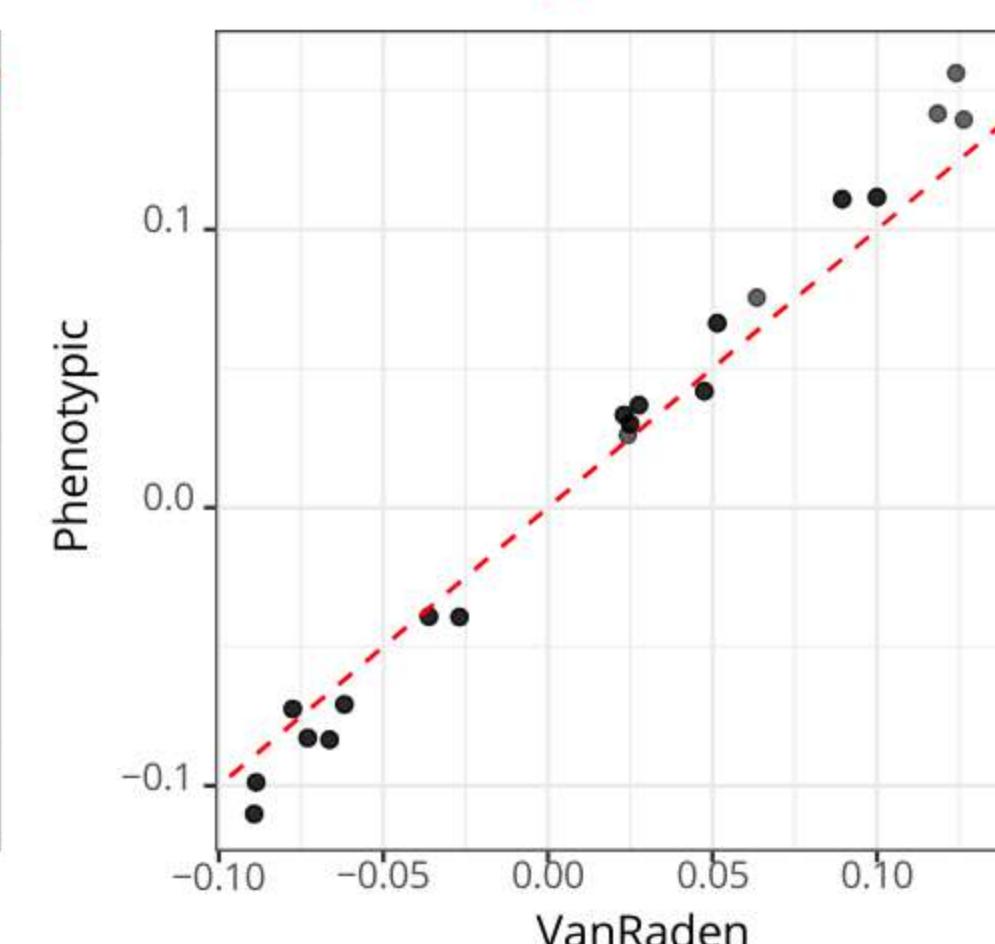
A6 VS Phenotypic

**B**

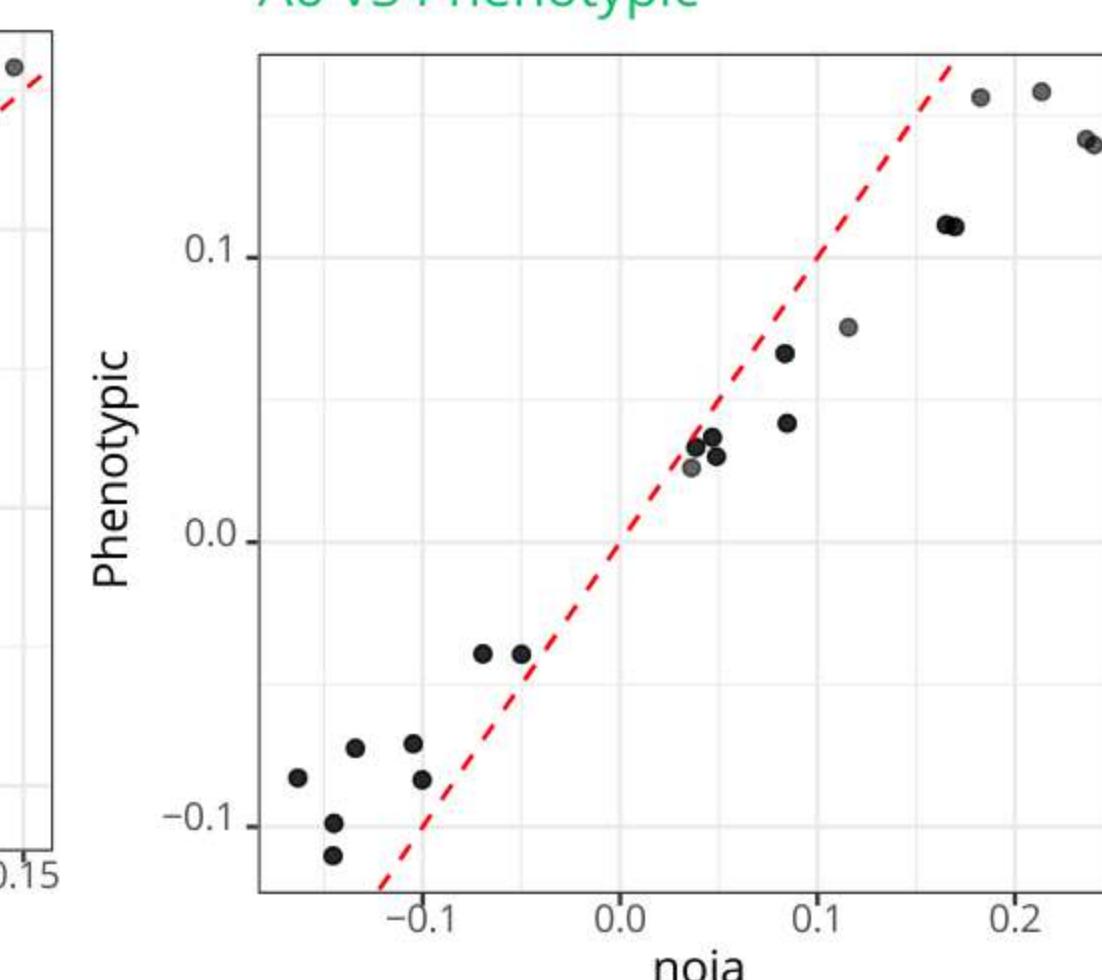
A6 VS Phenotypic

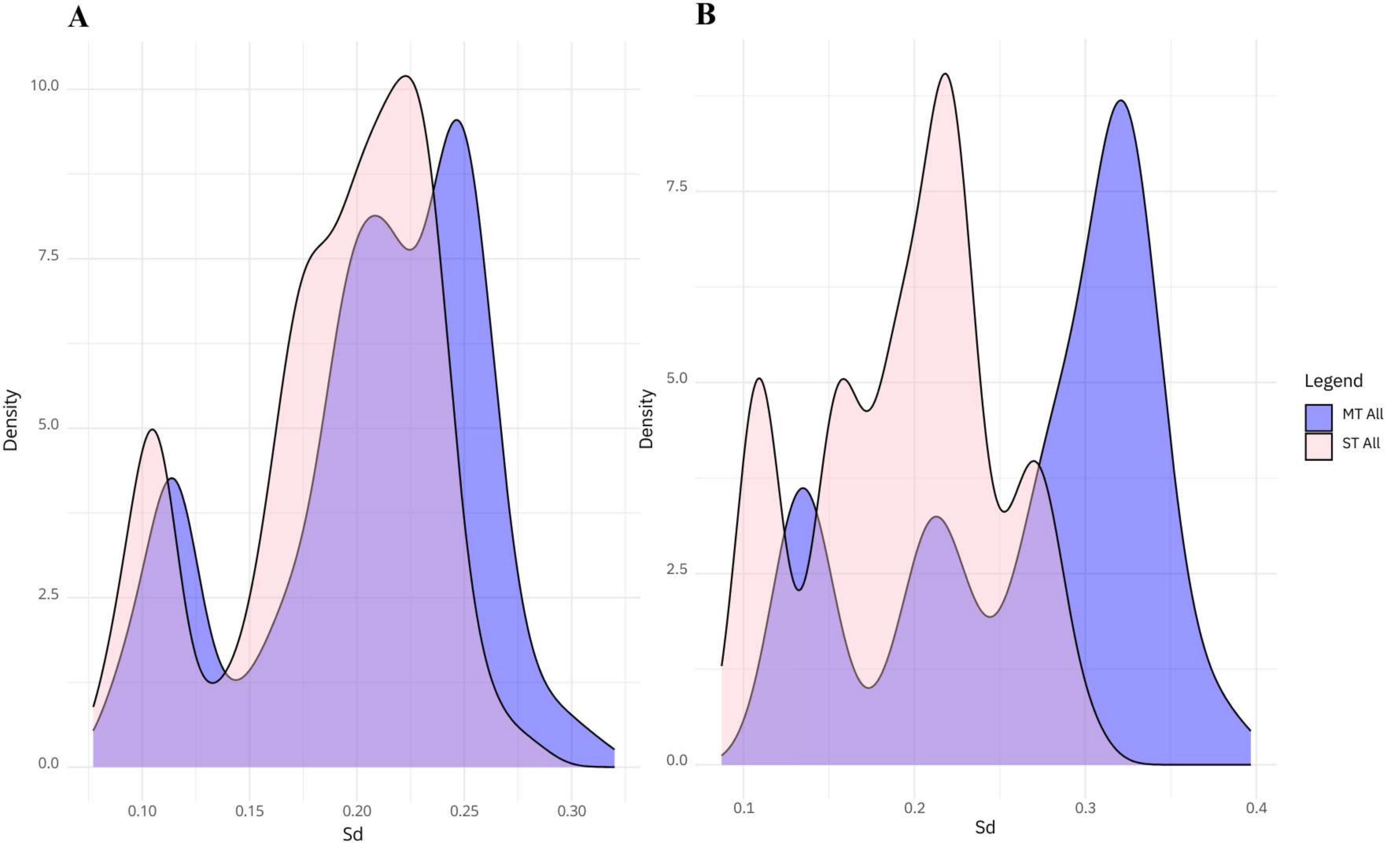


A6 VS Phenotypic

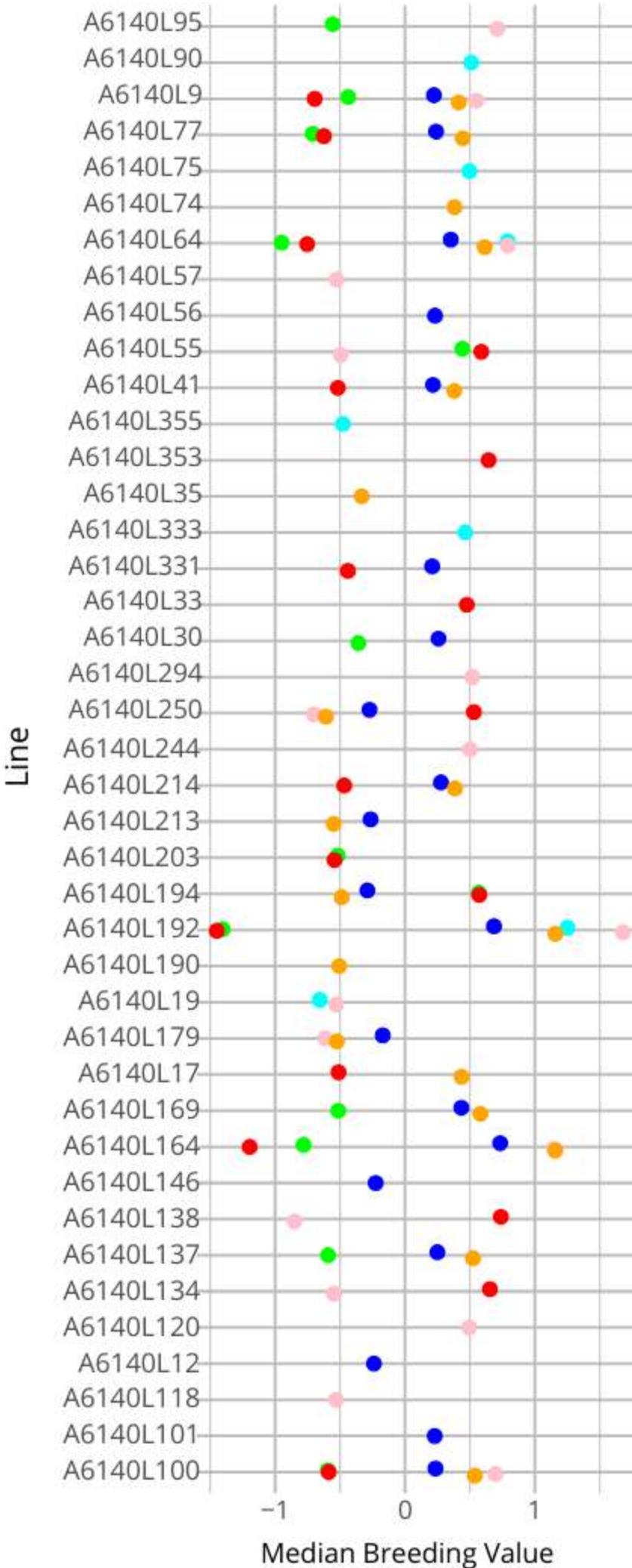


A6 VS Phenotypic

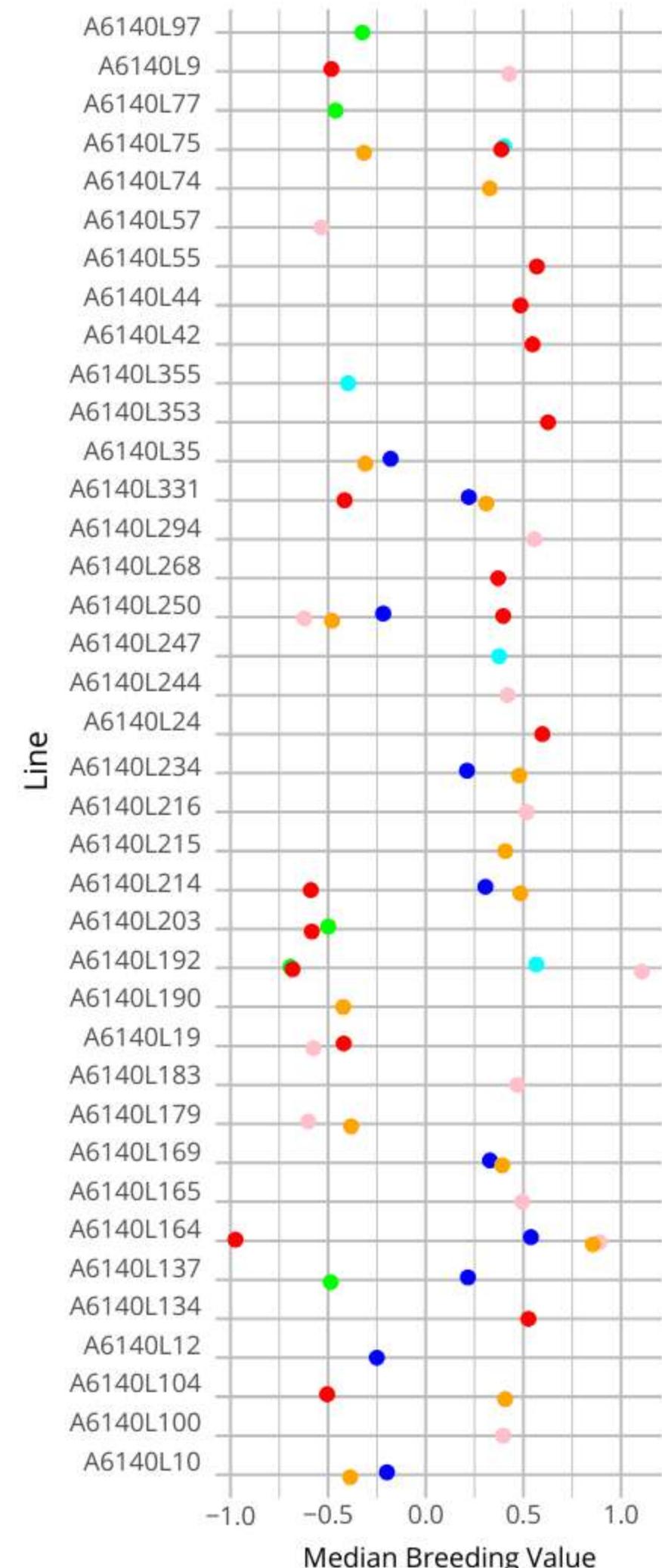




Multivariate Analysis

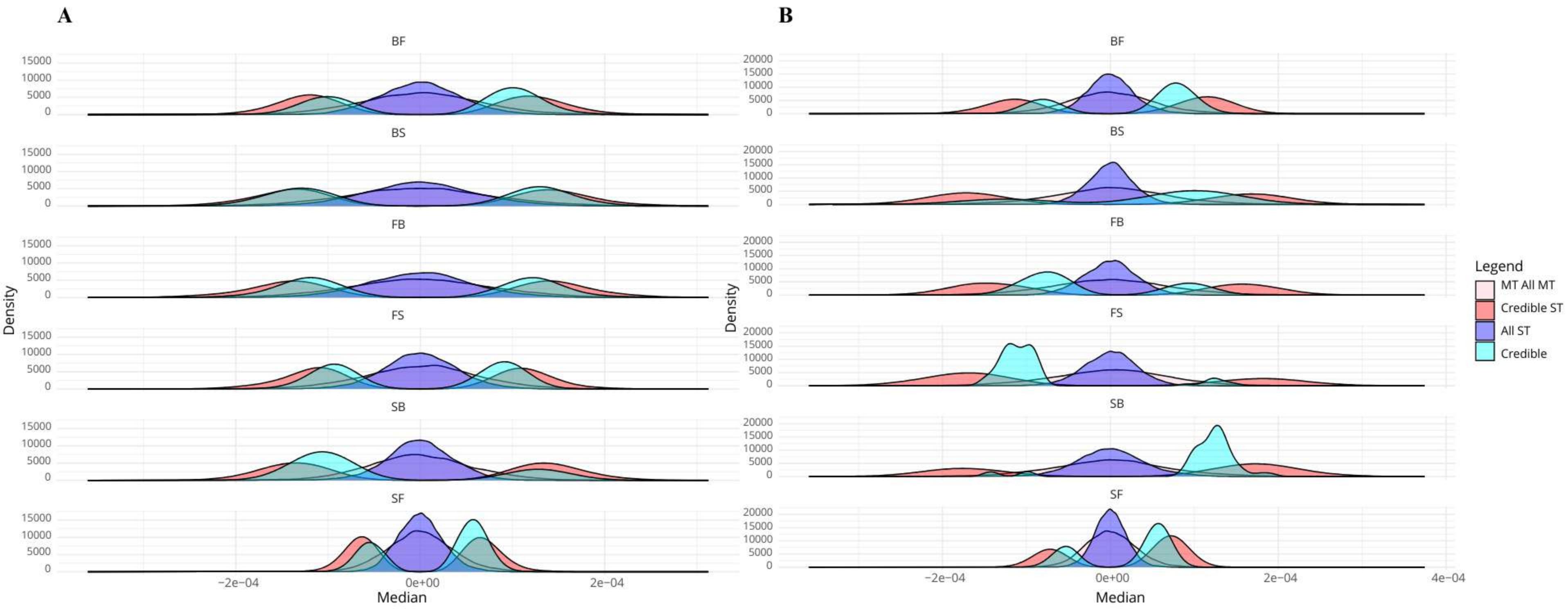


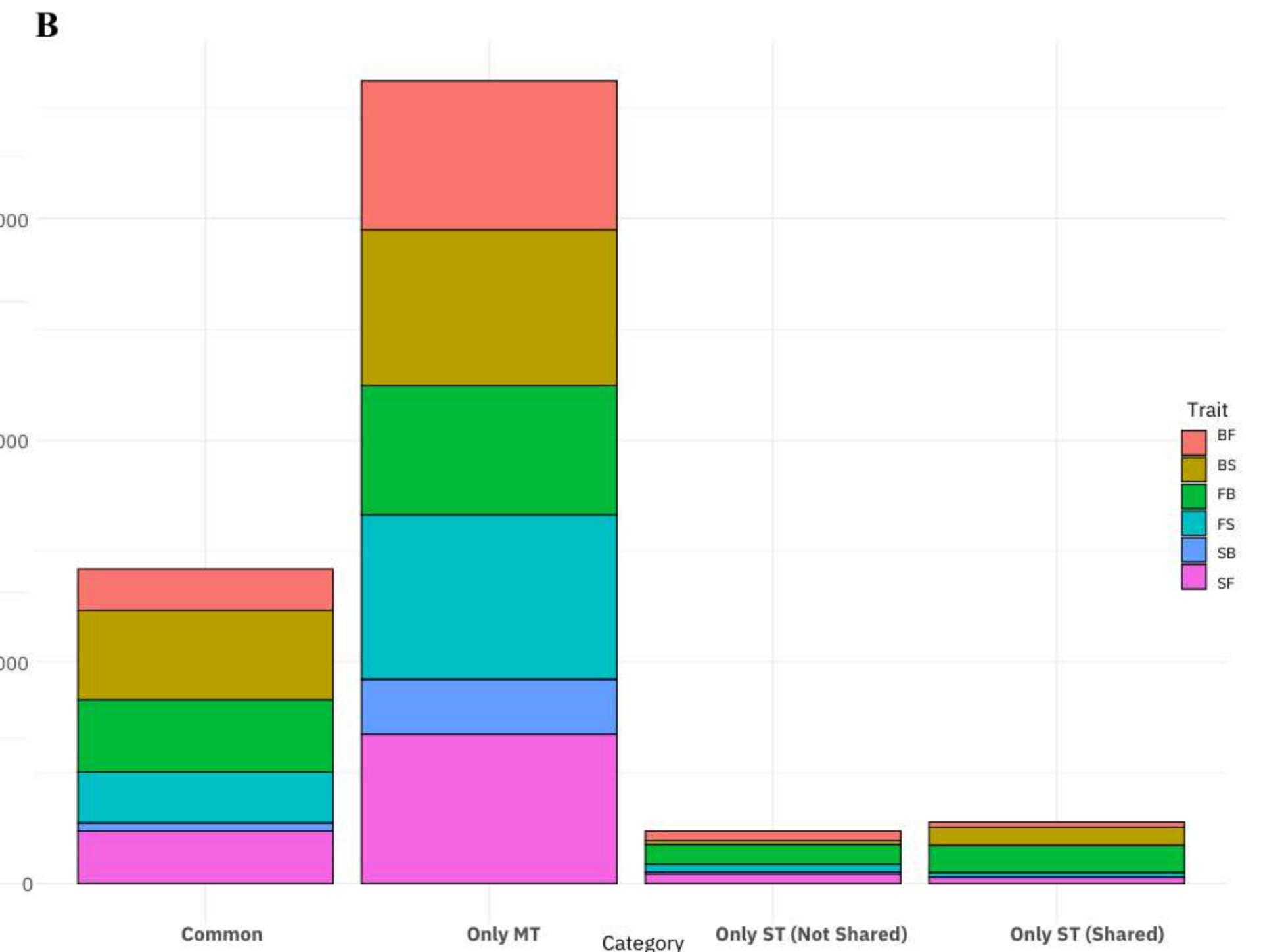
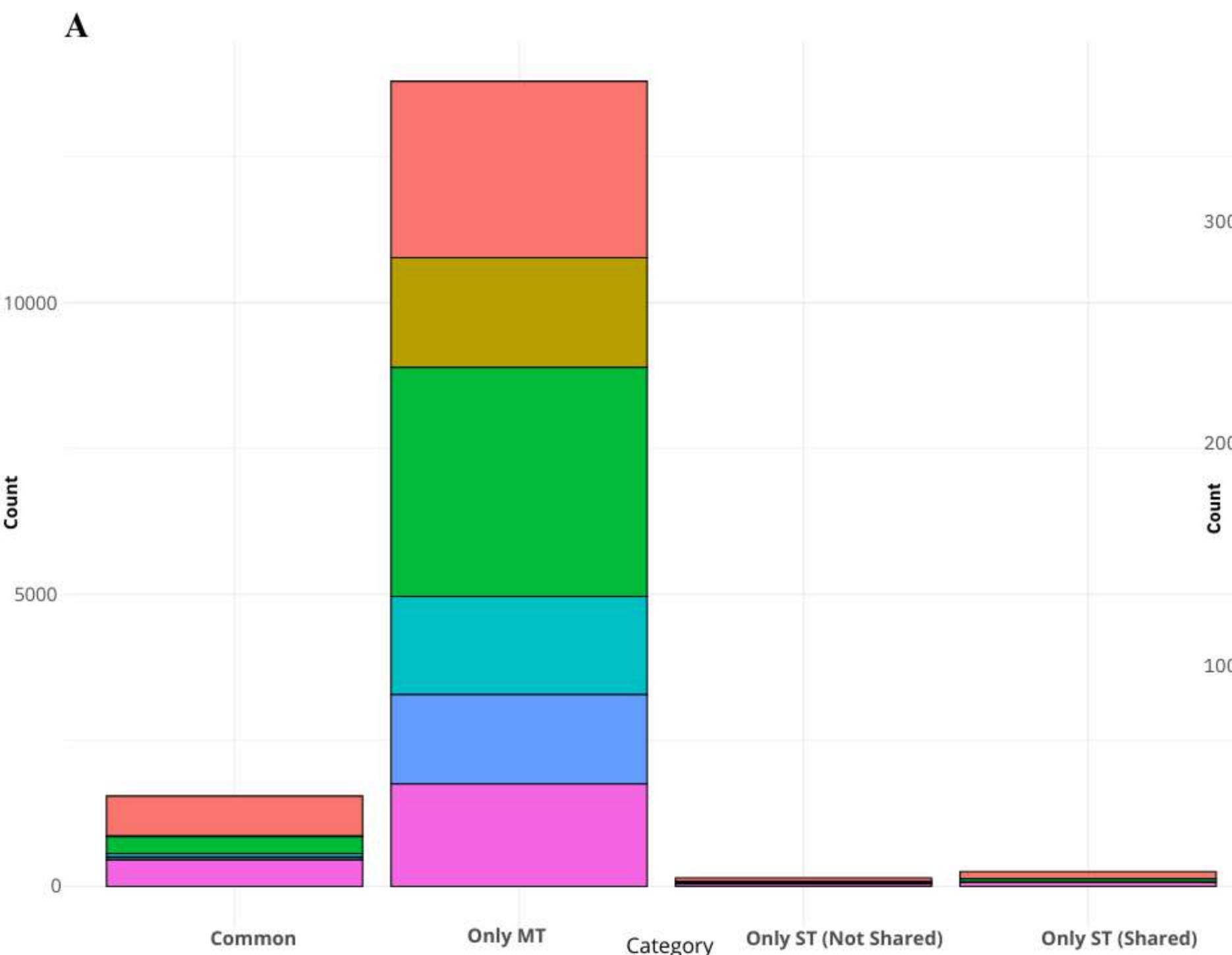
Univariate Analysis

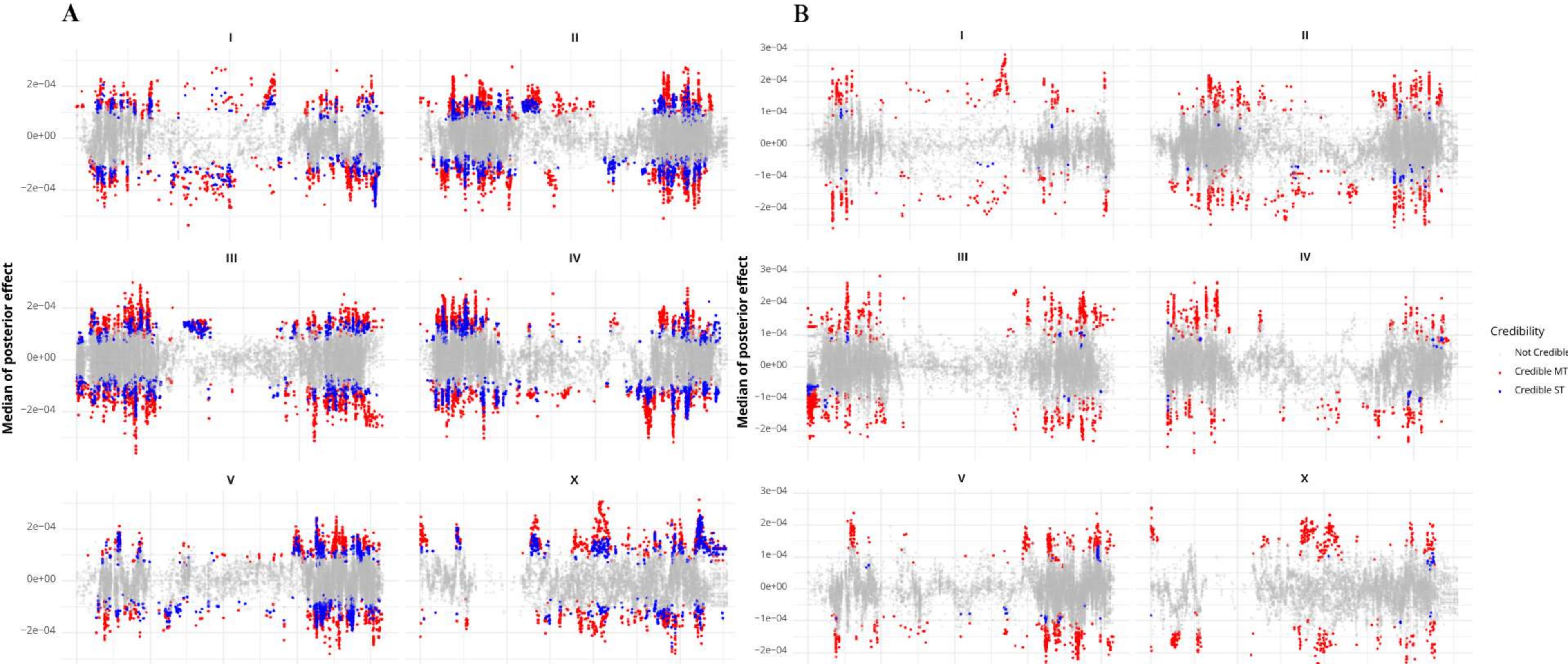


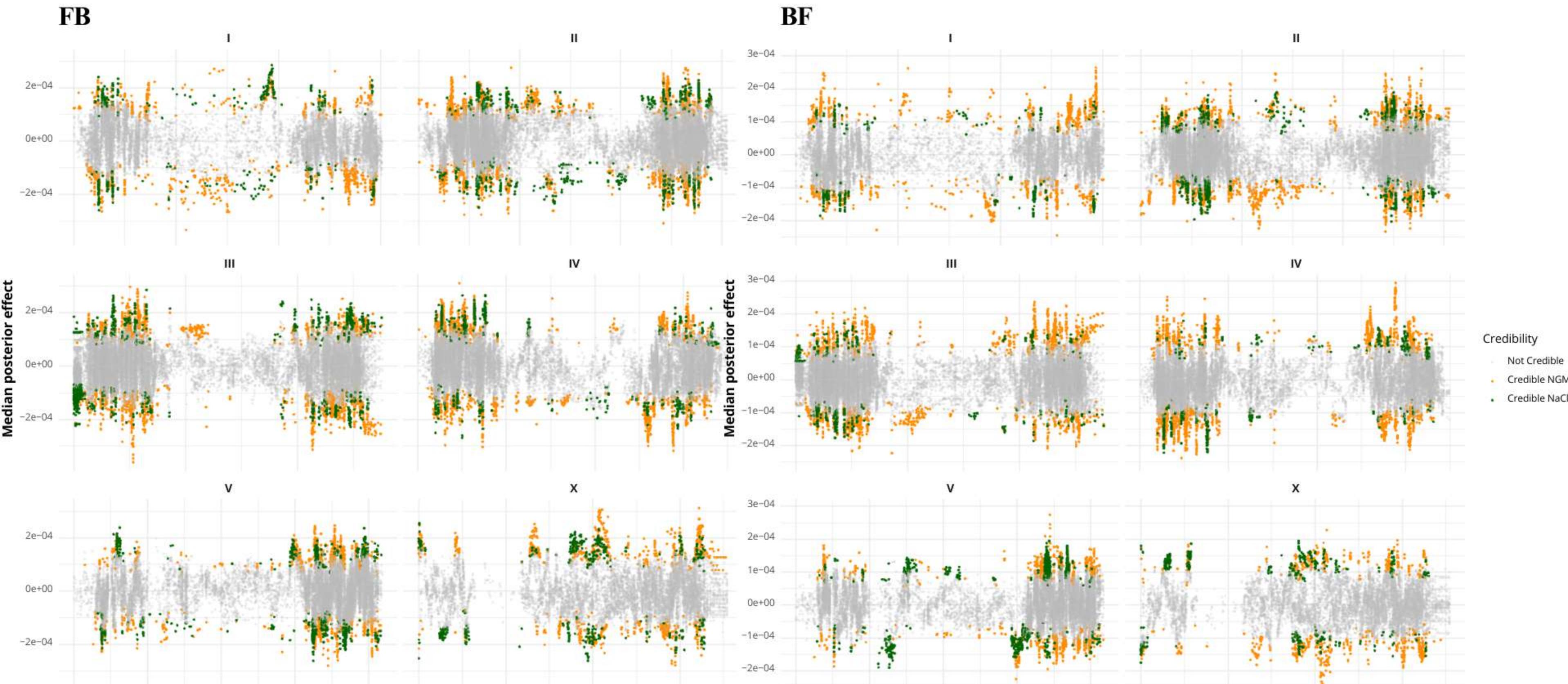
Trait

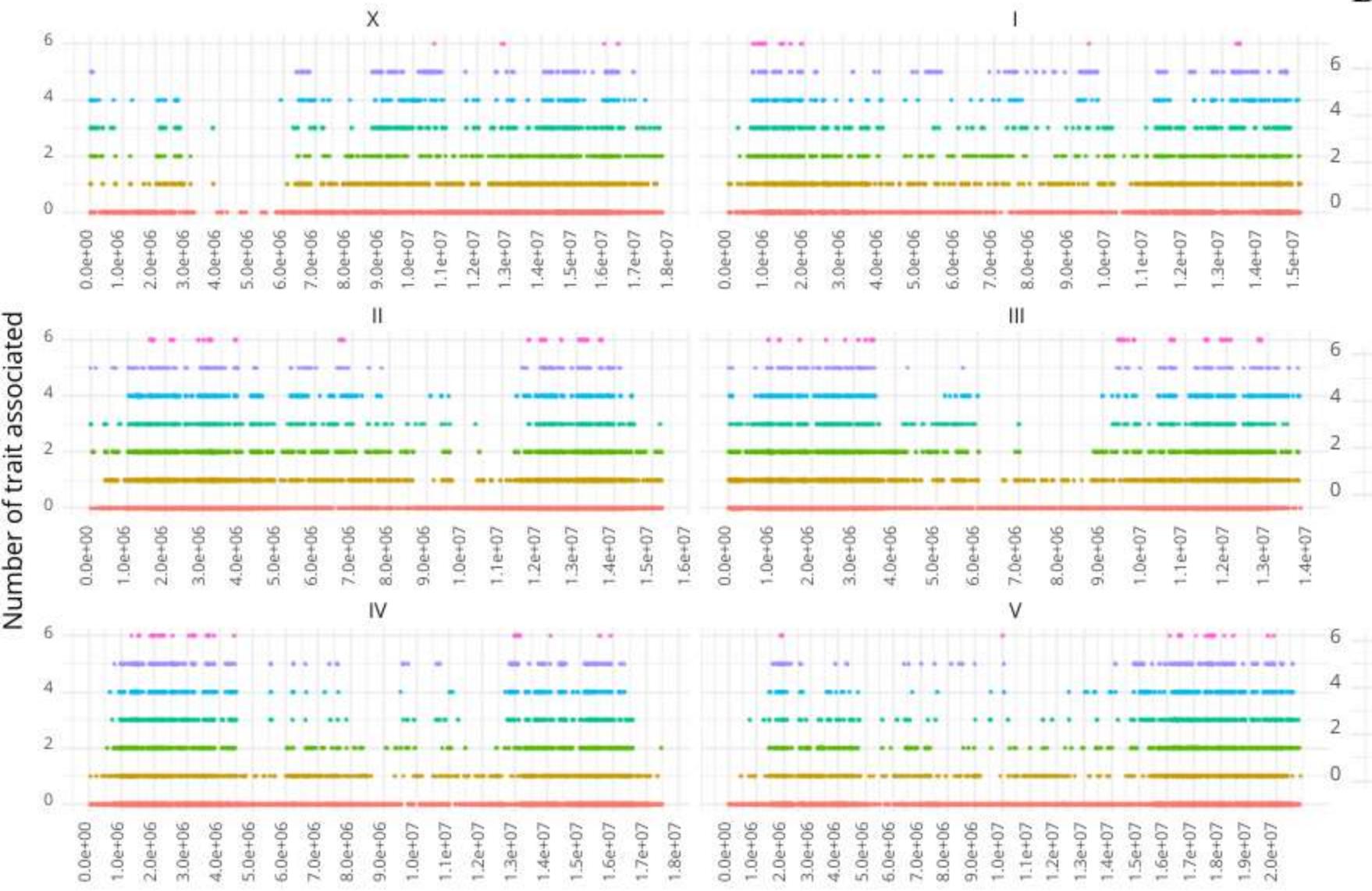
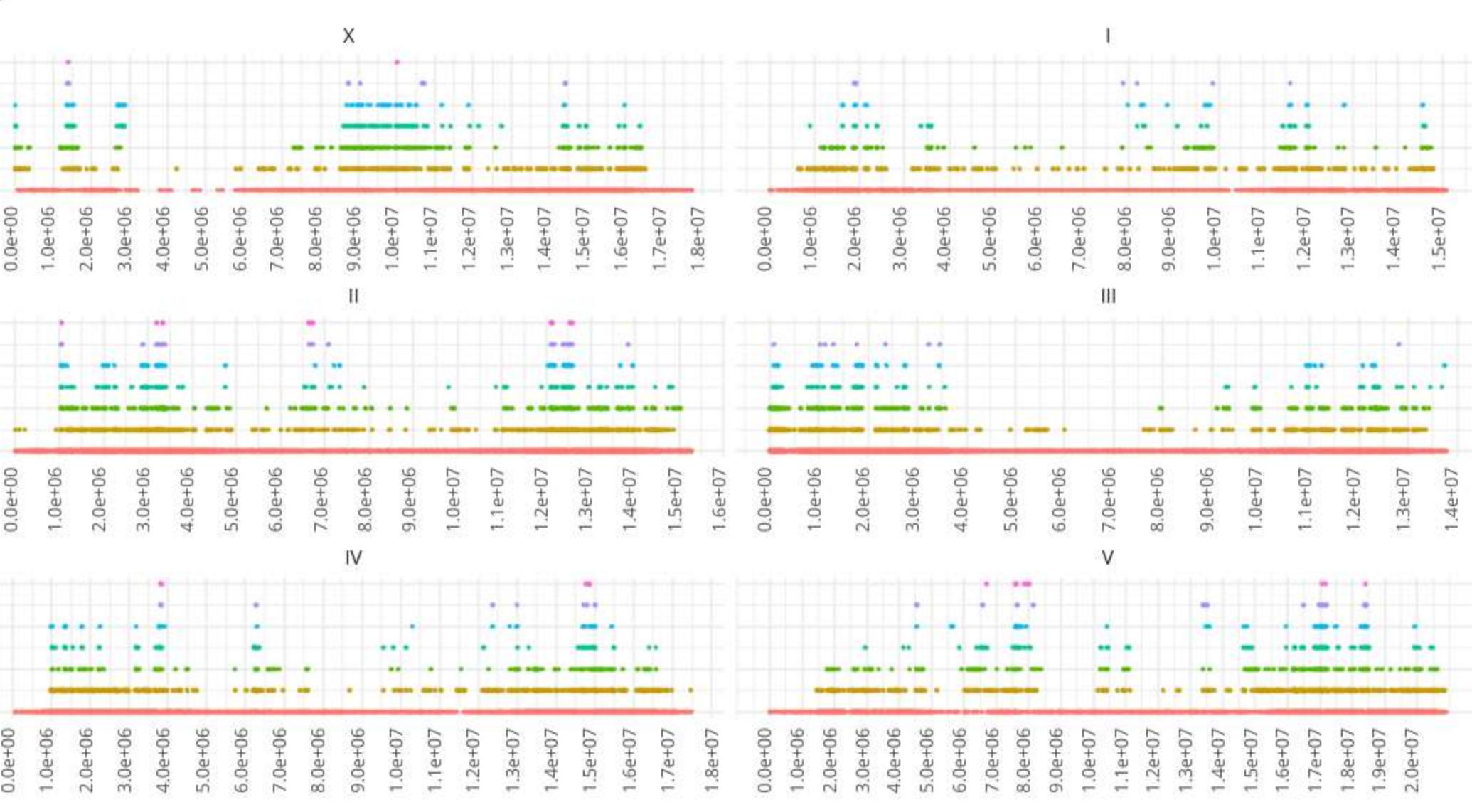
- BF
- BS
- FB
- FS
- SB
- SF



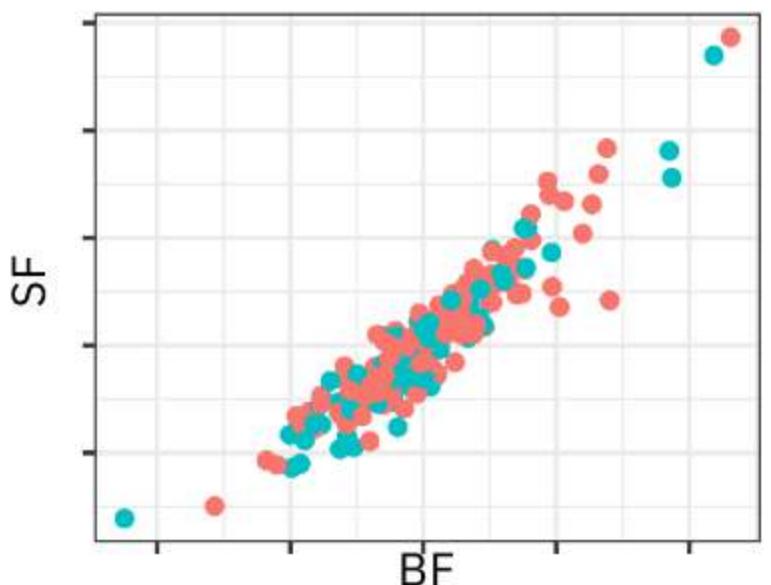




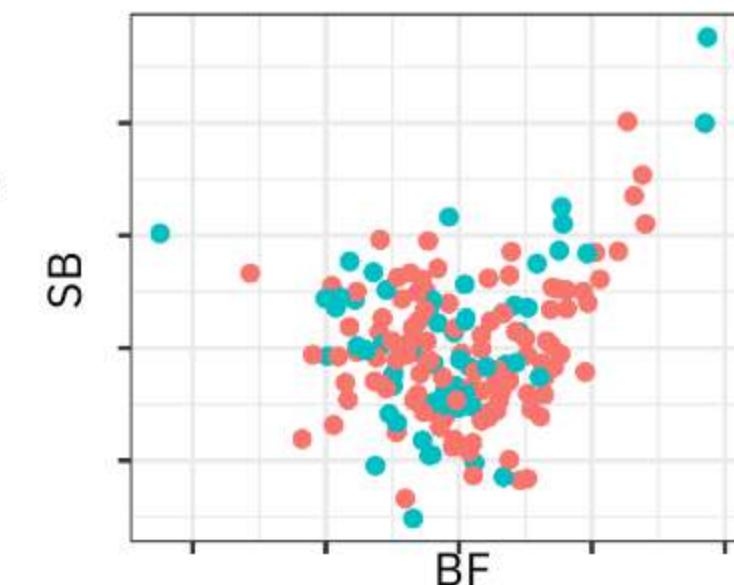


A**B**

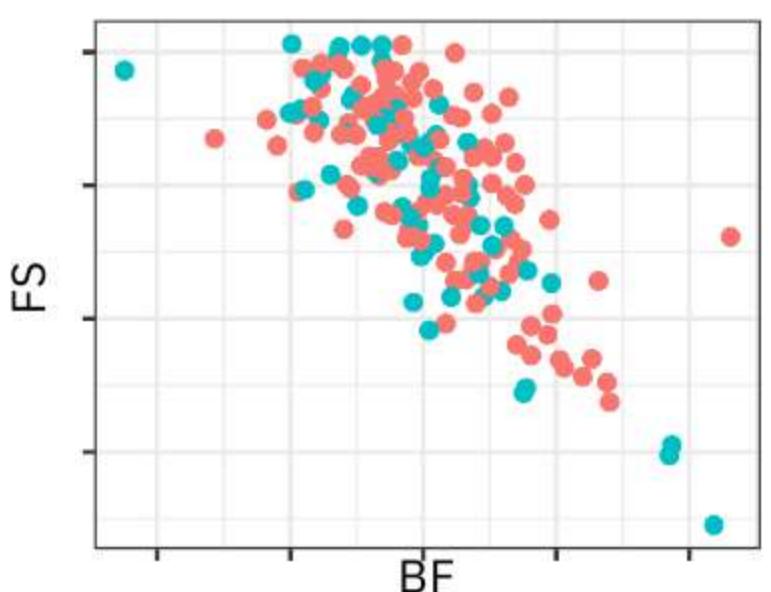
BF vs SF



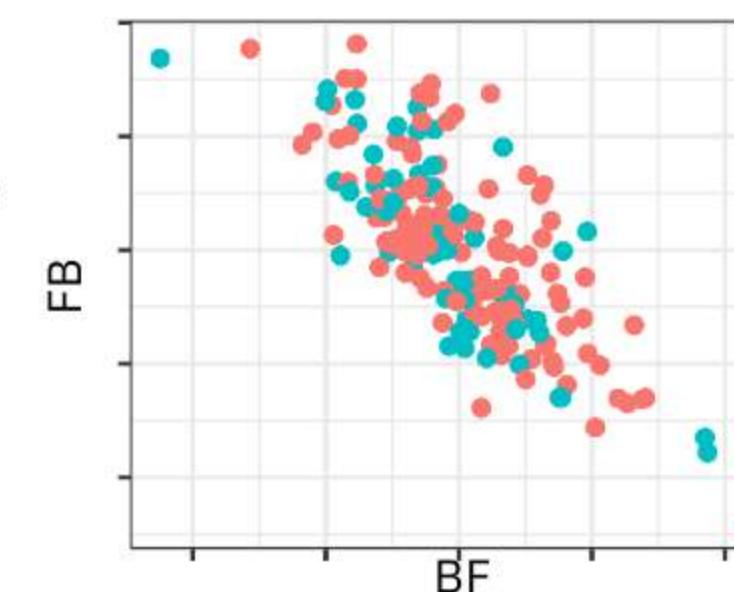
BF vs SB



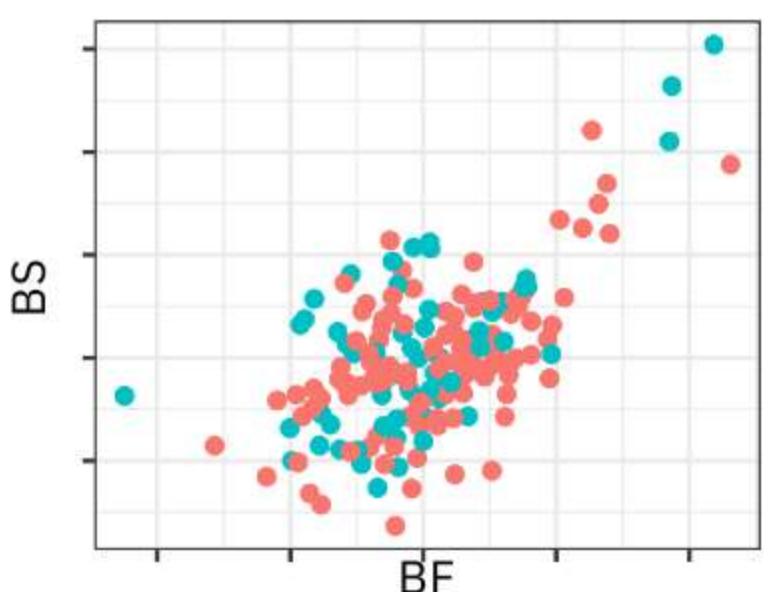
BF vs FS



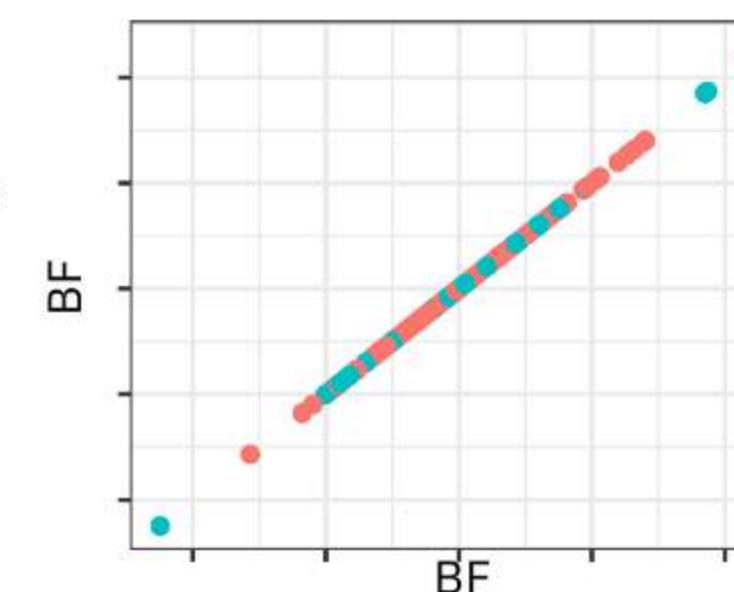
BF vs FB



BF vs BS



BF vs BF



Genotype

- 0
- 2

