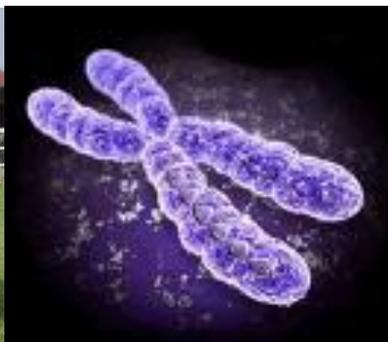


Interpretation génétique des ACP inter-groupes sur les haplotypes d'SNP

Application à la détection de région sous sélection
Denis Laloe et Matthieu Gautier



ALIMENTATION
AGRICULTURE
ENVIRONNEMENT



ACP

- Méthode standard et efficace pour explorer des gros fichiers de génotypage
- Ce papier propose une interprétation supplémentaire des résultats d'ACP pour la caractérisation de la structure génétique des populations
 - ✓ Et en particulier, comment les valeurs typologiques des SNP issus de l'ACP peuvent conduire à l'identification d'empreinte de sélection
 - ✓ Les variables d'entrée des ACP sont des haplotypes
 - ✓ Matrice $n \times m$ avec n haplotype et m SNP
 - ✓ Les SNP sont standardisés
 - ✓ Avec ces notations, la corrélation entre colonne i et j représente le LD entre les 2 SNP considérés
 - ✓ La matrice symétrique (ZZ'/n avec Z la matrice de SNP standardisée) est une matrice de LD

ACP

- Puisque les SNP sont des variables binaires, une analyse ACP sur la matrice haplotypique est aussi une analyse de correspondance multiple (ACM)
- Cette méthode est une référence pour l'analyse des tables de contingence multidimensionnelle
- Les ACP résument la variabilité totale parmi les individus sans différentier la variation intra et inter groupes
- Pour évaluer la relation entre différents clusters ou groupes (par exemple différentes races d'une espèce donnée), il faut isoler la variabilité inter groupe tout en faisant abstraction de la variabilité intra-groupe

Interpretation des scores de SNP

- Le score d'un SNP sur un axe est la corrélation du SNP avec cet axe
- En ACM, le score au carré pour un SNP sur un axe représente le % de variance du score haplotypique expliqué par ce SNP
- Plus ce score est élevé, plus les haplotypes sont distants
- Ce score correspond à une valeur typologique (contribution du SNP)

ACP inter groupes

- L'utilisation de groupes au sein d'une ACP permet de prendre en compte une structure au sein des individus
- A partir de Z (matrice des SNP standardisés), on construit une matrice Z^+ avec
 - En colonne, les SNP
 - En ligne, les différents groupes
 - La matrice contient, la valeur moyenne de chaque SNP pour chaque groupe
 - Z^+ est pondérée par la fréquence relative des groupes
- A partir de ces données, on calcul la meilleure combinaison linéaire (qui maximise la variance inter-groupes)

ACP intra groupes

- L'idée avec l'ACP intra-groupe est de s'affranchir des effets liés à la structure
- Pour cela, on génère la matrice Z- obtenue en centrant les données sur la moyenne par groupe.
- La variabilité intra-groupe peut être évaluée par la variance totale des scores individuels, sommé sur l'ensemble des composantes principales au sein du groupe
- $V = V_b + V_w$
 - V_b = somme des valeurs propres des ACP inter-groupes
 - V_w = somme des valeurs propres des ACP intra-groupe
 - V_b/V = contribution en terme de variance de la différenciation des individus au sein des groupes
 - Ce ratio est une mesure du F_{ST} .

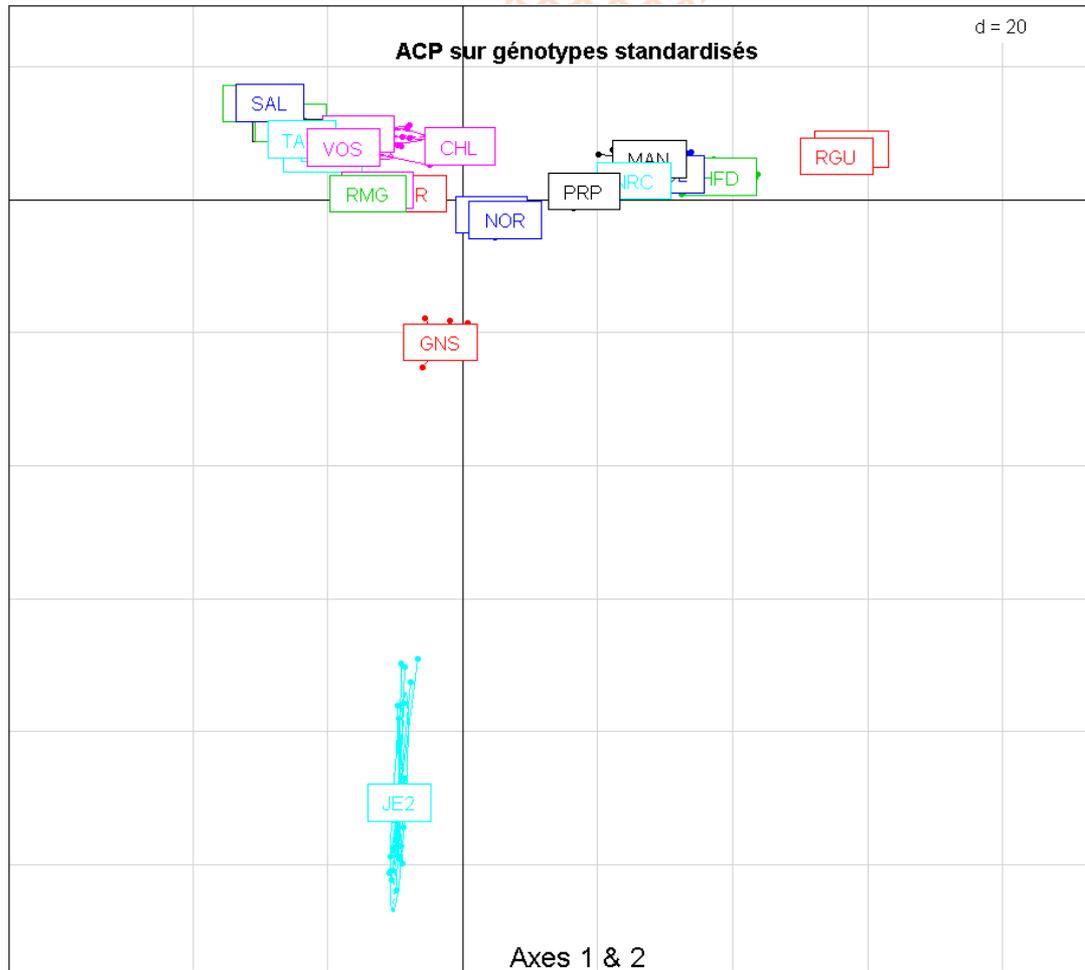


Applications sur les données bovines

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

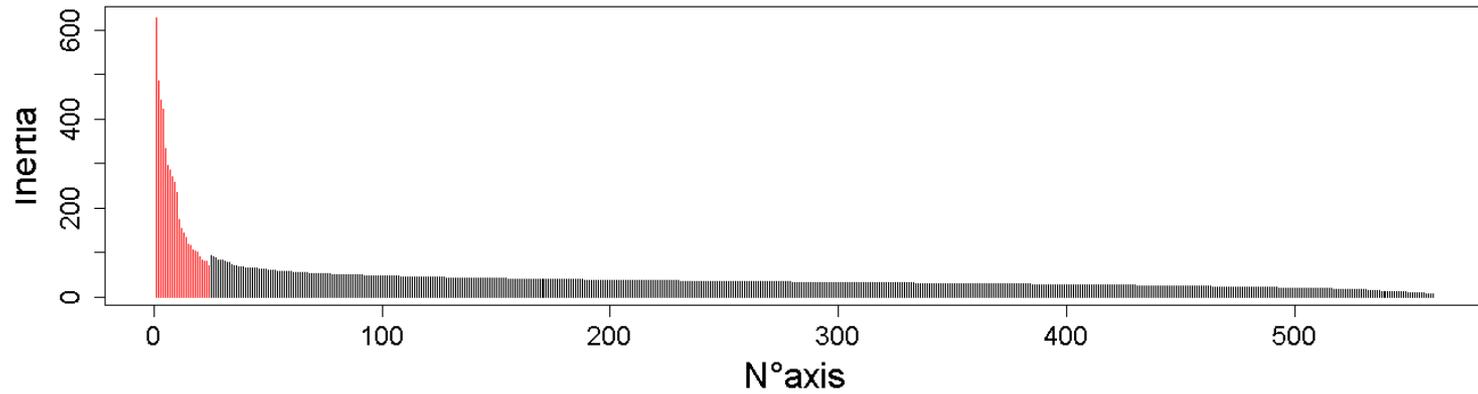
INRA

ACP

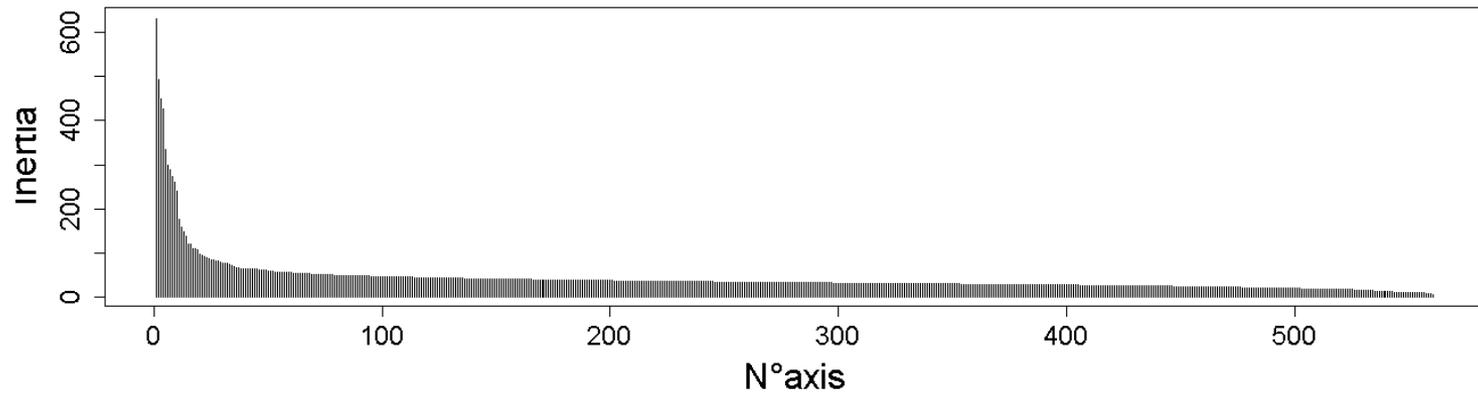


Variance par axe

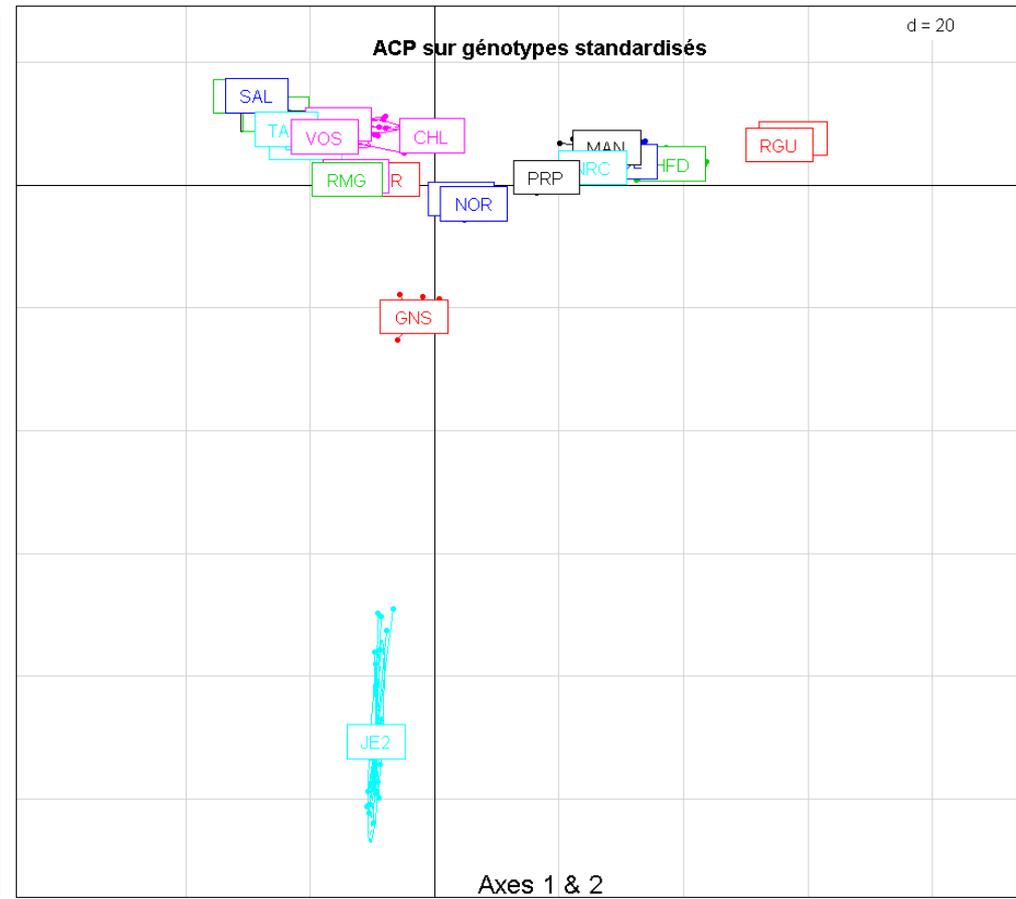
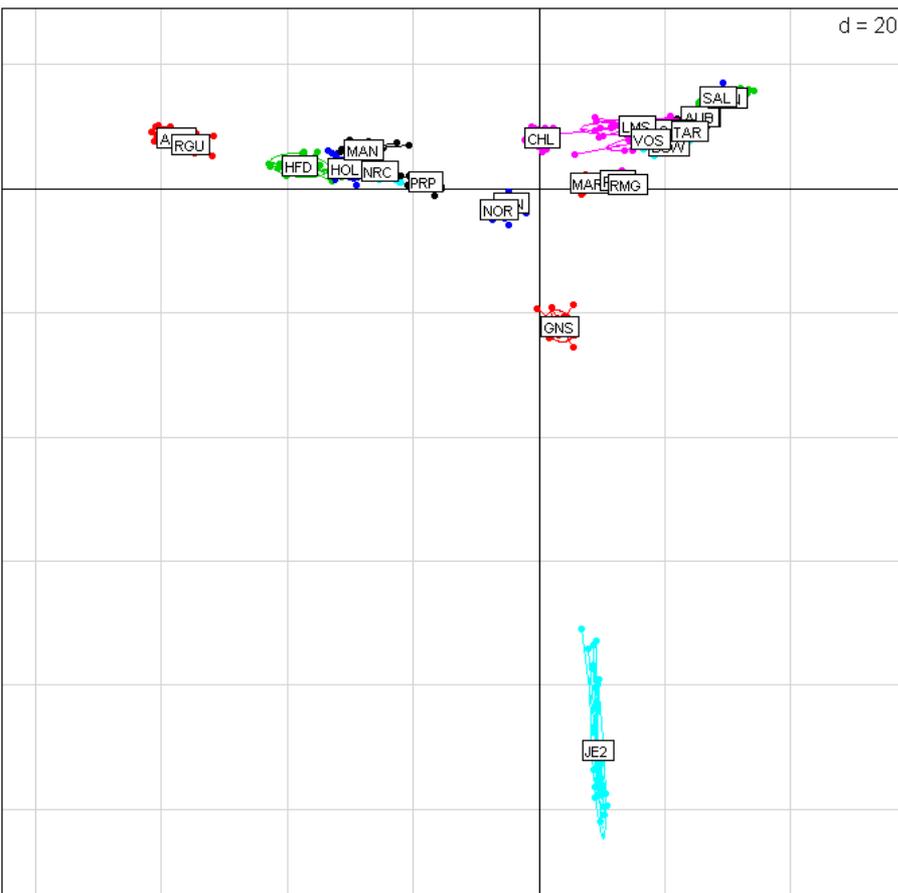
Valeurs propres des ACP inter + intra



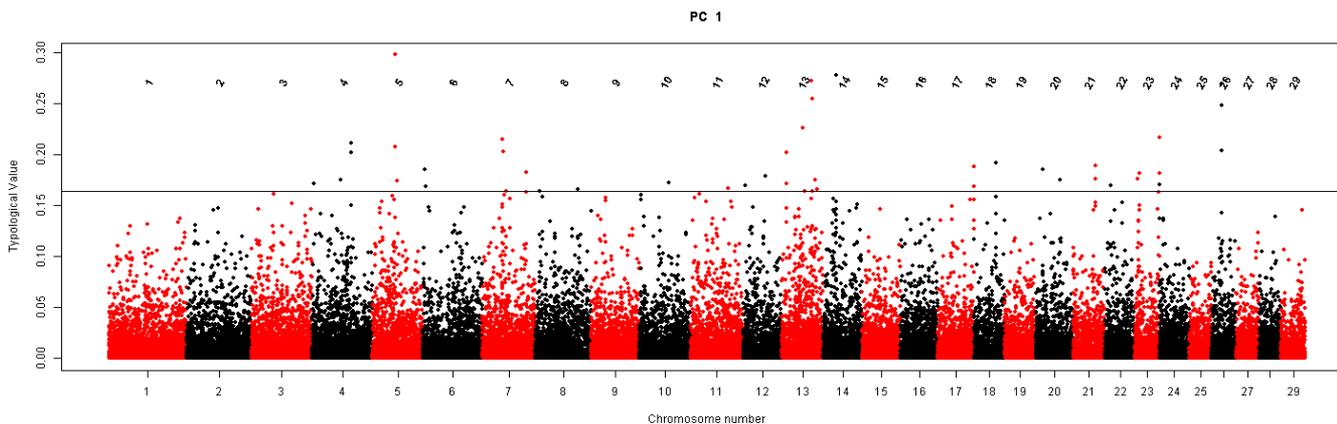
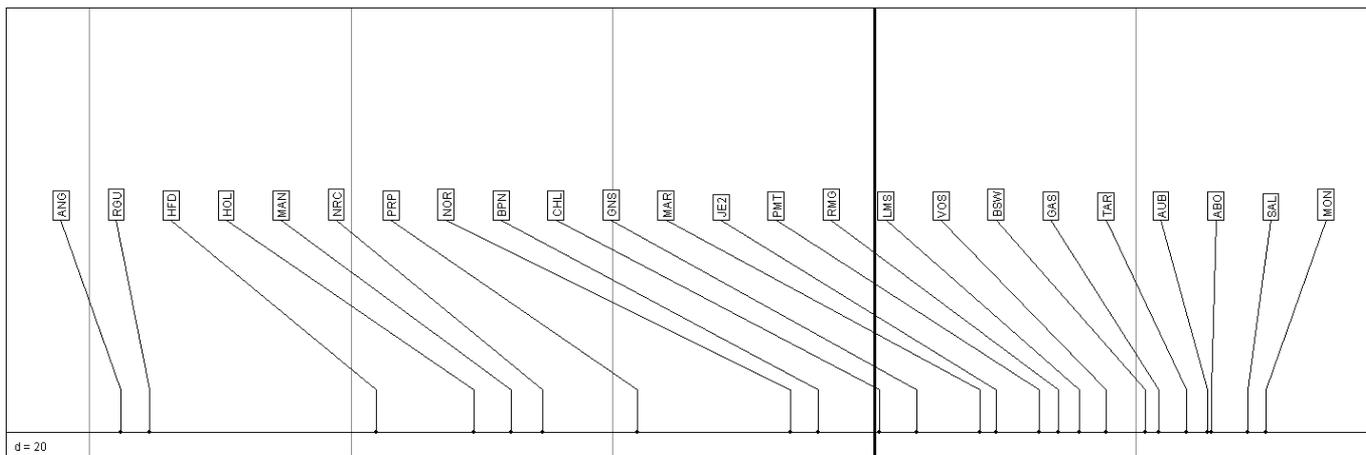
Valeurs propres de l'ACP sur les individus



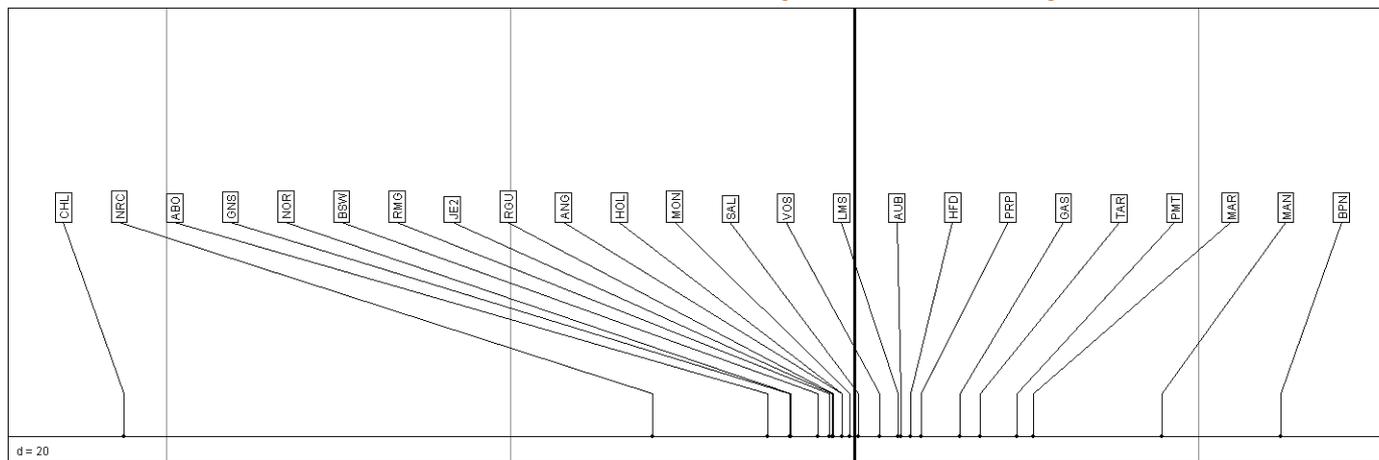
ACP inter-races



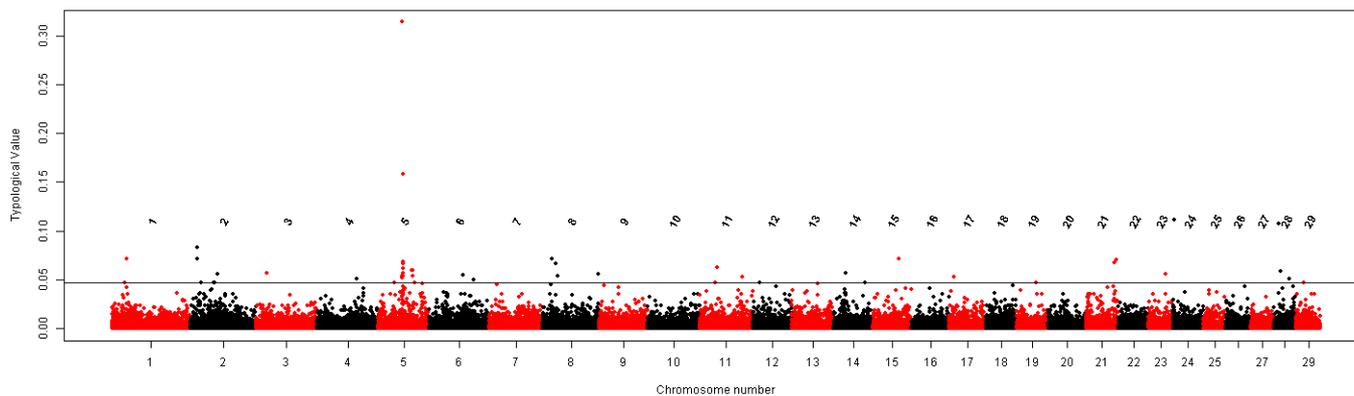
Contribution des SNP à la différenciation des races (axe1)



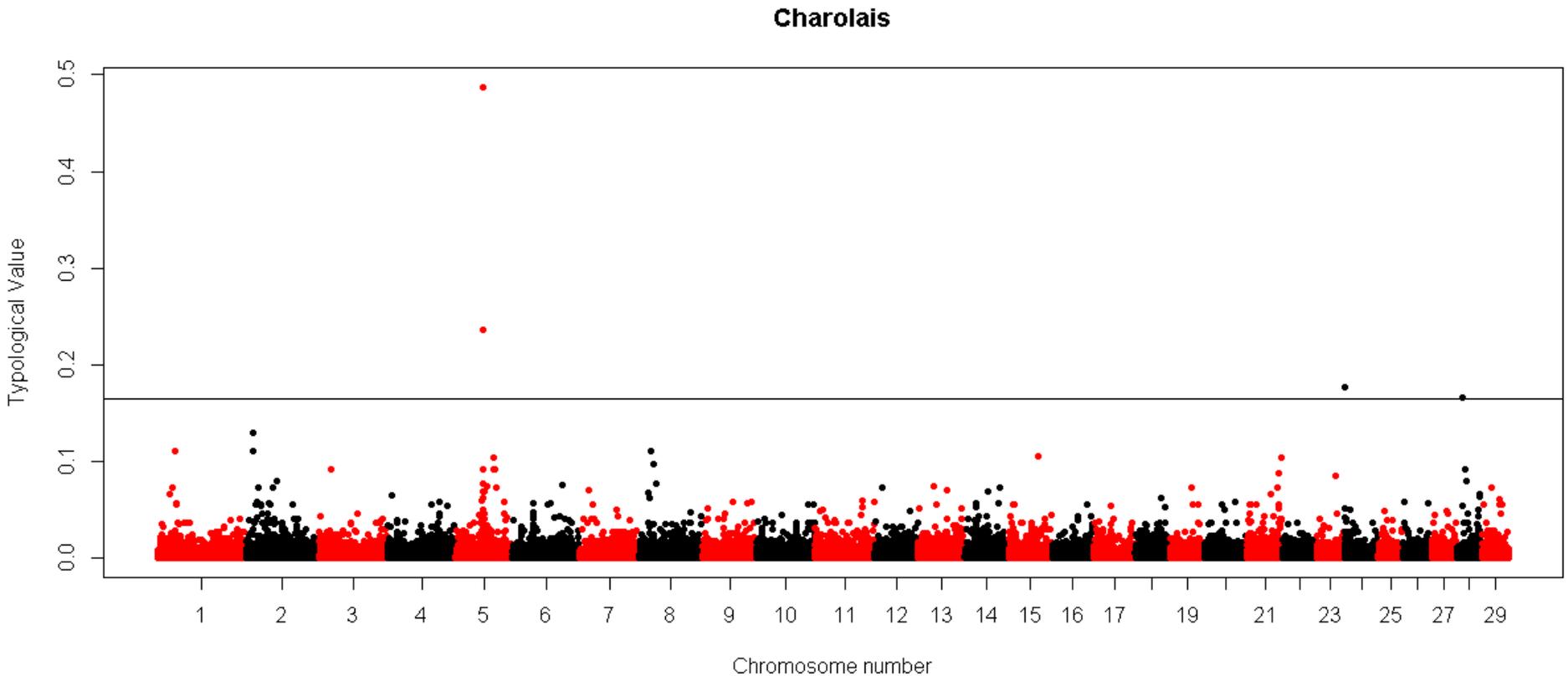
Contribution des SNP à la différenciation des races (axe13)



PC 13



Contribution des SNP à la race charolaise



ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA