

# Diversité génomique de porcs issus d'un croisement Large White x Piétrain

Audrey GANTEIL (1,2), Myriam COTTEREAU (2), Silvia RODRIGUEZ-RAMILO (1), Bruno LIGONESCHE (2), Catherine LARZUL (1)

(1) GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France

(2) SAS NUCLEUS, 35650 Le Rheu, France

[audrey.ganteil@inrae.fr](mailto:audrey.ganteil@inrae.fr)

## Diversité génomique de porcs issus d'un croisement Large White x Piétrain

Un des bénéfices attendus du croisement entre lignées, pratique répandue dans la filière porcine, peut être l'accroissement de variabilité génétique, notamment dans un contexte de création de populations synthétiques. Cette augmentation de la variabilité, liée à l'état hétérozygote des individus, reste difficile à préciser lorsque seule l'information généalogique est disponible. Il est dorénavant possible de caractériser la variabilité génétique présente dans les populations porcines à une échelle génomique via l'analyse de données de génotypage. Une des approches possibles est la détection de régions homozygotes le long du génome, qui permet de quantifier la consanguinité génomique chez un individu. La quantification de ces régions homozygotes pouvant être réalisée à l'échelle du génome ou du chromosome, elle permet de caractériser à la fois la consanguinité génomique globale d'un individu ou une consanguinité locale en des zones plus ciblées du génome. Nous proposons dans cette étude d'utiliser l'information génomique afin d'estimer la consanguinité via la détection de régions homozygotes présentes dans les génomes de reproducteurs de races pures Large White et Piétrain et de leurs produits croisés. L'ensemble des individus a été génotypé sur la puce porcine IlluminaSNP60. Comme attendu, un nombre moins important de segments homozygotes est détecté chez les animaux croisés par rapport aux deux populations parentales. De plus, la longueur totale de ces segments est plus faible chez les animaux croisés. Même si ces observations peuvent être interprétées comme une diminution de la consanguinité, la conservation de segments homozygotes chez les descendants croisés signifie que certains haplotypes sont partagés entre les populations parentales.

## Genomic diversity of pigs from a Large White x Pietrain crossbreeding

One of the expected benefits of crossbreeding, a common practice in the pig industry, may be the increase in genetic variability, especially in a context of creation of synthetic populations. This increase in variability, related to the heterozygous status of individuals, remains difficult to specify when only genealogical information is available. It is now possible to characterize genetic variability in pig populations at a genome scale by analyzing genotypic data. One possible approach is to detect homozygous regions along the genome to quantify genomic consanguinity in an individual. These homozygous regions can be quantified at the genome or chromosome scale. It is then possible to characterize both overall genomic consanguinity of an individual or a local one in specific areas of the genome. In this study, we used genomic information to estimate inbreeding by detecting homozygous regions in the genomes of purebred Large White dams and Pietrain sires and their crossbred offspring. All individuals were genotyped using the IlluminaSNP60 porcine chip. As expected, fewer homozygous segments were detected in crossbred animals than in their parental populations. In addition, the total length of these segments was lower in crossbred animals. These observations can be interpreted as a decrease in consanguinity; however, the conservation of homozygous segments in crossbred offspring means that some haplotypes were shared between parental populations.

## INTRODUCTION

La création de progrès génétique à long terme nécessite d'être vigilant sur la gestion de la variabilité génétique des populations, afin de préserver le potentiel d'évolution et d'adaptation des animaux. Aujourd'hui, grâce à l'information génomique, il est possible de caractériser la diversité génétique des animaux de manière plus précise qu'à partir d'analyses de la seule information généalogique (VanRaden, 2008).

La diversité génomique peut être appréhendée par l'estimation de la consanguinité génomique. Par définition, la consanguinité au sens génétique du terme est la probabilité qu'un individu, à un locus donné, possède deux allèles identiques par descendance, provenant donc d'un ancêtre commun à ses deux parents. La consanguinité génomique peut être estimée par la quantification des segments homozygotes présents dans le génome (ROH pour « Runs of Homozygosity »). Ces ROH résultent de l'existence d'un ancêtre commun aux parents de l'individu, dont certains fragments chromosomiques se retrouvent à l'état homozygote chez l'individu, traduisant un état consanguin. L'apparition de ROH au sein d'une population est influencée par la présence d'éventuels goulets d'étranglement génétiques, d'accouplements consanguins ou suite à une sélection intensive (Peripolli *et al*, 2017).

En élevage porcin, le croisement est largement utilisé pour la production d'animaux destinés à l'abattoir afin d'exploiter la complémentarité entre races et de profiter des phénomènes d'hétérosis. Un troisième bénéfice du croisement, notamment lors de la constitution d'une nouvelle lignée, est l'accroissement de la variabilité génétique avec l'hypothèse de générer de l'hétérozygotie. Cependant, cette augmentation de variabilité reste difficile à prévoir lorsque seule l'information généalogique est disponible (Bidanel, 1992). Afin de caractériser cette évolution de variabilité d'un point de vue génomique, nous avons réalisé une détection de ROH chez des individus de races pures, Large White et Piétrain et chez leurs produits croisés.

## 1. MATERIEL ET METHODES

### 1.1. Génotypages et contrôle qualité des données

Les données génomiques utilisées pour cette étude ont été obtenues auprès de l'entreprise de sélection NUCLEUS. Un total de 80 verrats Piétrain, 240 femelles Large White et 441 croisées Piétrain x Large White ont été génotypés par le laboratoire de Gènes Diffusion avec la puce porcine IlluminaSNP60. Un contrôle qualité a été réalisé pour ces données avec le logiciel PLINK version 1.9 (Purcell *et al*, 2007). Seuls les marqueurs présents sur les autosomes ont été conservés pour l'analyse. Il a été vérifié que l'ensemble des animaux possédait plus de 90% de marqueurs informatifs. Les marqueurs ayant plus de 5% de génotypes manquants dans la population ont été écartés. A l'issue de ce contrôle qualité, la totalité des animaux et 56 695 marqueurs ont été conservés. Ce jeu de données a été utilisé sans filtre supplémentaire dans l'ensemble des analyses de cette étude.

### 1.2. Analyse de la diversité génomique

#### 1.2.1. Distances génétiques entre les individus

Un positionnement multidimensionnel (MDS) a été réalisé avec le logiciel PLINK 1.9. Cette analyse permet de visualiser les distances génomiques entre les animaux considérés.

Pour chaque paire d'individus, les distances sont par défaut exprimées en comptage de marqueurs en commun entre les deux individus. Seuls les résultats des deux premières dimensions ont été conservés.

#### 1.2.2. Taille efficace des populations parentales

La taille efficace d'une population est un paramètre génétique qui permet d'estimer la dérive génétique dans une population. Elle correspond à un effectif d'une population idéale (sans sélection et avec des accouplements au hasard) qui présenterait le même accroissement de consanguinité que la population réelle. Ainsi, la taille efficace permet d'évaluer le nombre de génomes indépendants présents dans une population.

La taille efficace des populations Large White et Piétrain a été calculée jusqu'à 32 générations afin d'estimer la dérive génétique présente dans ces deux populations parentales. L'analyse a été réalisée avec le logiciel SNeP (Barbato *et al*, 2015) avec les paramètres par défaut hormis le paramètre de fréquence de l'allèle mineur (MAF), qui n'a pas été pris en compte ici.

#### 1.2.3. Détection de régions homozygotes

La détection de ROH a été réalisée avec le logiciel PLINK 1.9. Ce dernier utilise une fenêtre glissante dont la taille est fixée en nombre de SNP. Elle se déplace de SNP en SNP et à chaque position, la caractéristique « homozygote » ou non de la fenêtre est déterminée en tenant compte des paramètres fixés au préalable. Pour cette analyse, il est possible de distinguer les paramètres relatifs à la fenêtre glissante puis ceux relatifs aux ROH. Nous avons choisi de fixer la taille de la fenêtre glissante à 30 SNP et nous avons autorisé un SNP non génotypé par fenêtre. Aucun SNP hétérozygote n'est autorisé par fenêtre afin d'obtenir des ROH strictement homozygotes à l'issue de l'analyse. Le ratio du nombre de fenêtres « homozygotes » sur le nombre total de fenêtres est calculé pour chaque SNP. Si ce dernier est supérieur à 5%, alors le SNP est considéré comme homozygote. Si au moins 30 SNP consécutifs sont homozygotes, un ROH peut être alors défini dans la région génomique étudiée à condition que la région respecte également les caractéristiques suivantes : une longueur minimale de 1000 Kb et une densité minimale en SNP d'un SNP pour 100 Kb.

L'ensemble des marqueurs ayant passé le contrôle qualité précédemment réalisé a été conservé pour la détection de ROH, afin d'être informatif, notamment dans les régions du génome ne présentant que peu de variabilité. L'estimation des ROH le long du génome permet de calculer un coefficient de consanguinité génomique noté  $F_{ROH}$  dont la formule est la suivante :

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{AUTOSOMES}}$$

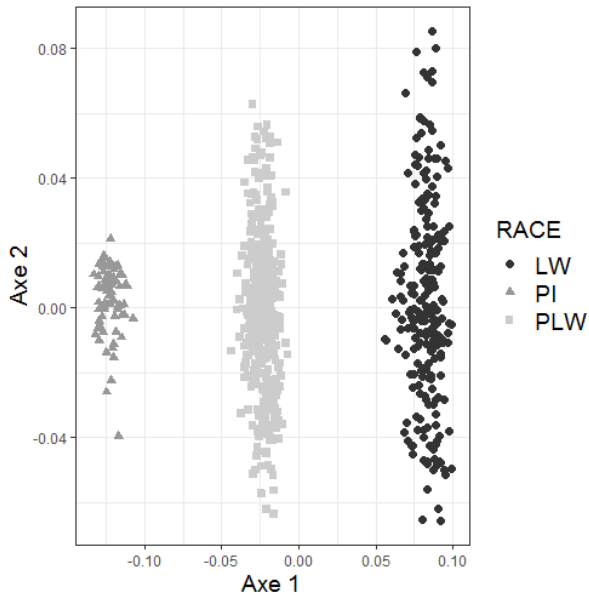
$\sum L_{ROH}$ , est la longueur totale de l'ensemble des ROH détecté chez un individu et  $L_{AUTOSOMES}$  la longueur totale des autosomes (2263711134 pb). Sur le même principe, un  $F_{ROH}$  peut être calculé à l'échelle de chaque chromosome de l'individu.

Ce coefficient de consanguinité génomique a été comparé avec le coefficient de consanguinité génétique  $F_X$  estimé à partir de données pedigree. A partir de la généalogie des animaux sur 10 générations, nous avons utilisé le logiciel Pedig afin de calculer le coefficient de consanguinité d'après la méthode de Meuwissen et Luo (Boichard, 2002).

## 2. RESULTATS ET DISCUSSION

### 2.1. Distances génétiques et tailles efficaces des populations parentales

L'analyse MDS permet de visualiser la population en trois sous-groupes distincts, correspondant aux deux races étudiées et à leur croisement (Figure 1).



**Figure 1** - Distances génétiques entre les individus des trois types génétiques.

LW : Large White, PI : Piétrain ; PLW : Piétrain x Large White

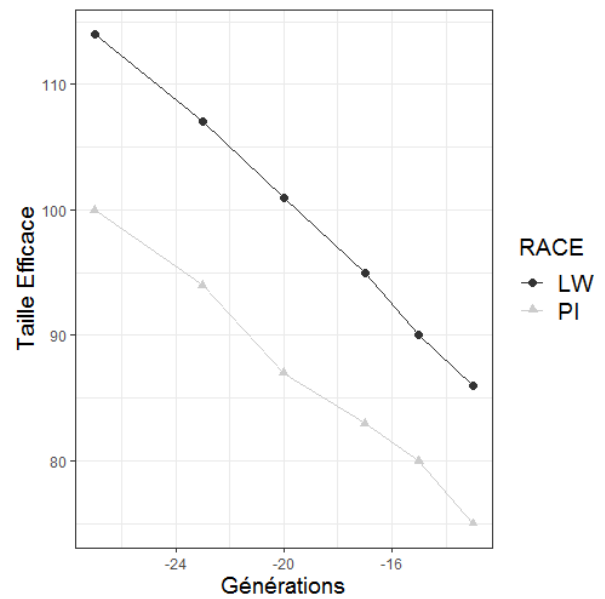
Le premier axe permet de différencier les individus selon la race et le second met en évidence la variabilité génétique intra-race. Le groupe constitué par les animaux croisés, Piétrain x Large White, se situe à mi-distance entre les deux populations parentales par rapport à l'Axe 1. Ce résultat illustre bien la composition chromosomique, moitié Piétrain et moitié Large White des animaux croisés. Concernant la diversité intra-race pour les deux populations parentales, les animaux Piétrain ont tendance à être plus regroupés contrairement aux animaux Large White qui se répartissent de manière plus dispersée sur l'Axe 2.

L'analyse des tailles efficaces des populations parentales permet de visualiser une décroissance progressive de la taille efficace des populations et par conséquent de la diversité génétique dans les populations de Piétrain et Large White étudiées (Figure 2). Sur 32 générations, la taille efficace de la population Piétrain étudiée demeure toujours inférieure à celle de la population Large White. Il y a 32 générations, la taille efficace est estimée à 123 pour les Large White et 108 pour les Piétrain. La taille efficace la plus récente (13 générations) estimée est de 86 pour les Large White et 75 pour les Piétrain.

Les résultats de la Figure 1 et 2 semblent indiquer une plus faible variabilité génomique chez les Piétrain par rapport aux Large White. Une explication à ce résultat peut être la présence d'effets de type « goulot d'étranglement » dans l'histoire démographique des Piétrain français, cette race ayant presque disparue de France au début des années 1980.

Ces résultats préliminaires sont cependant à nuancer car les effectifs d'animaux génotypés sont différents entre les deux populations et l'effectif d'animaux Piétrain est beaucoup plus faible que celui des Large White (respectivement 80 animaux

contre 240). Une caractérisation plus approfondie de la diversité génomique est donc pertinente ici afin de compléter ces premiers résultats.



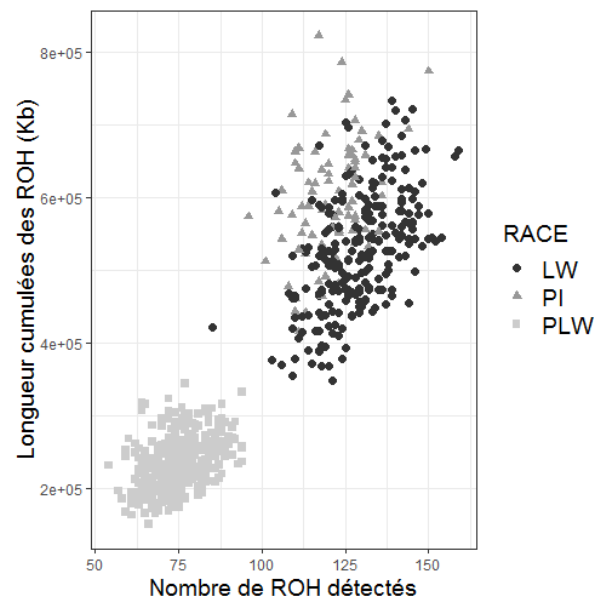
**Figure 2** - Taille efficace des populations parentales sur 32 générations.

LW : Large White, PI : Piétrain

### 2.2. Consanguinité génomique

#### 2.2.1. A l'échelle du génome

La détection de ROH permet ici de caractériser la proportion d'homozygotie présente dans les génomes de chaque individu (Figure 3).



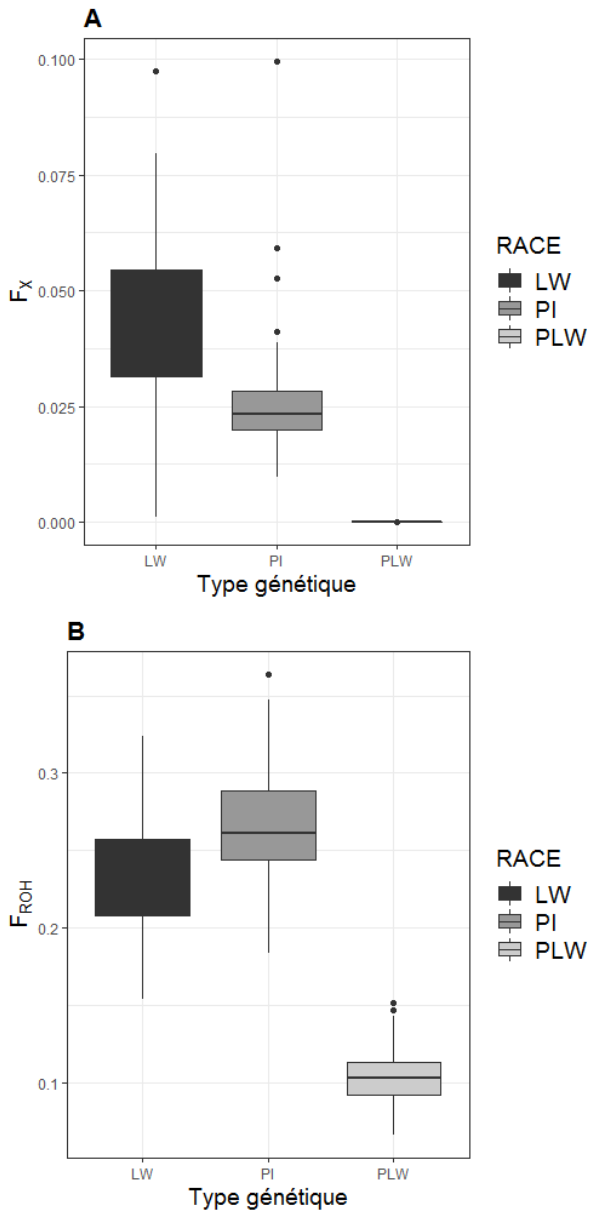
**Figure 3** - Longueur cumulée des ROH détectés en fonction du nombre de ROH détectés chez chaque individu. Un individu est représenté par un point.

LW : Large White, PI : Piétrain ; PLW : Piétrain x Large White

Les animaux de race pure, Large White et Piétrain, présentent à la fois un nombre de ROH plus important et une longueur cumulée de ROH plus élevée que les animaux croisés. Ce résultat illustre ici la création d'hétérozygotie, liée au

croisement des deux races parentales, diminuant par conséquent le nombre et la longueur des ROH chez les individus croisés.

Le coefficient de consanguinité ROH,  $F_{ROH}$ , a été comparé au coefficient de consanguinité calculé à partir de données de pedigree  $F_X$  (Figure 4).



**Figure 4** - Coefficients de consanguinité génétique (A) et génomique  $F_{ROH}$  (B) estimés pour les trois types génétiques.

LW : Large White, PI : Piétrain ; PLW : Piétrain x Large White

Le coefficient  $F_X$  est plus élevé pour la population Large White que pour la population Piétrain et  $F_X$  est nul pour leurs descendants (par définition,  $F_X$  est nul pour des individus issus de parents non apparentés).

Contrairement à  $F_X$ ,  $F_{ROH}$  est légèrement supérieur chez les Piétrain par rapport aux Large White. Une explication possible à ce résultat est que  $F_X$  ne tient pas compte de la consanguinité éventuellement présente pour les fondateurs de la population, en les considérant comme non apparentés. Or, nous avons vu précédemment avec les tailles efficaces de population que celle des Piétrain était inférieure à celle des Large White (Figure 2). Cette diversité moindre des Piétrain n'est pas prise en compte

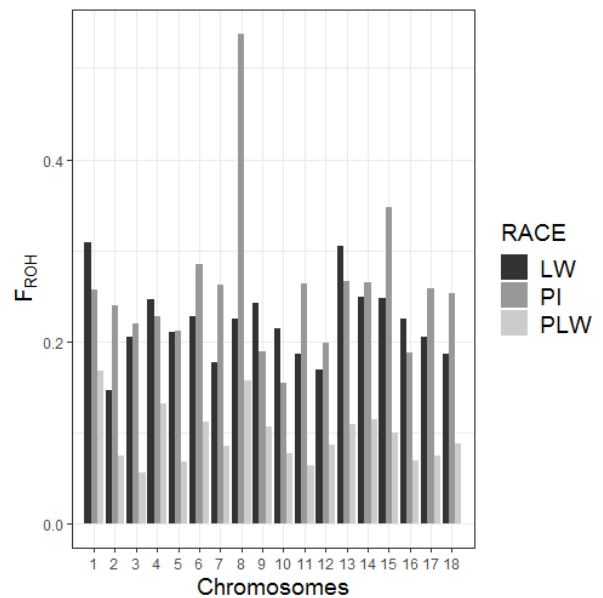
lors du calcul du coefficient de consanguinité sur données pedigree alors qu'elle l'est lors du calcul du  $F_{ROH}$ . Une seconde explication à ce résultat peut également être la différence de profondeur de pedigree entre les animaux Large White et Piétrain étudiés. En effet, la population Piétrain française est largement issue de Piétrain importés d'Allemagne et d'Autriche pour lesquels, les informations généalogiques sont limitées (Maignel et Labroue, 2001).

Pour la population croisée, il est possible ici d'estimer la consanguinité génomique. Le coefficient  $F_{ROH}$  obtenu pour les animaux croisés est plus faible que celui obtenu pour les deux populations parentales en raison du nombre moins élevé et de la longueur plus faible des ROH détectés (Figure 3).

Ainsi, ces coefficients de consanguinité doivent être interprétés indépendamment car ils ne décrivent pas le même type de variabilité.  $F_X$  traduit l'utilisation plus ou moins intensive de certains reproducteurs dans une population alors que  $F_{ROH}$  permet lui de quantifier chez chaque individu son degré d'homozygotie.

### 2.2.2. A l'échelle chromosomique

Nous avons déterminé le coefficient de consanguinité  $F_{ROH}$  pour chaque chromosome (Figure 5).



**Figure 5** - Coefficient de consanguinité  $F_{ROH}$  moyen estimé par chromosome chez les trois types génétiques.

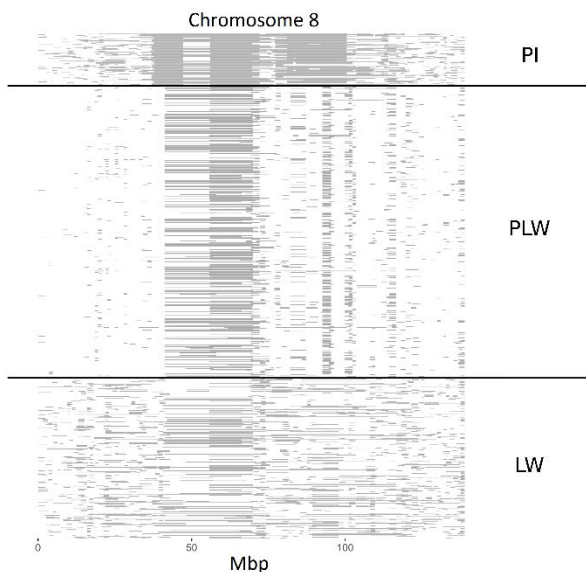
LW : Large White, PI : Piétrain ; PLW : Piétrain x Large White

Les deux races parentales présentent des  $F_{ROH}$  par chromosome plus élevés par rapport à la population croisée. Ce résultat est cohérent avec les précédents résultats de cette étude et illustre la génération d'homozygotie à une échelle locale. Les animaux Piétrain présentent un  $F_{ROH}$  plus élevé que les Large White pour tous les chromosomes, hormis les chromosomes 1, 4, 9, 10 et 13. Le chromosome 8 de la race Piétrain présente le  $F_{ROH}$  le plus élevé de l'analyse, avec une valeur moyenne de 0,53 contre 0,22 pour les Large White et 0,15 pour les animaux croisés. Ce résultat signifie qu'en moyenne, pour ces animaux, plus de la moitié du chromosome 8 est constituée de régions homozygotes.

Afin d'analyser plus en détail les ROH présents sur le chromosome 8, chaque segment homozygote a été représenté le long du chromosome pour les trois types génétiques (Figure 6).

Les ROH sont en moyenne d'une longueur de 5,27 Mb pour les Large White, de 5,71 Mb pour les croisés et de 10,75 Mb pour les Piétrain. Ce résultat montre à nouveau la conservation de ROH chez des animaux issus d'un croisement. Cela a déjà été observé en porc mais seulement pour un croisement où les génotypes des potentiels animaux croisés avaient été simulés par rapport aux génotypes des parents de race pure (Howard *et al*, 2016). La présence de ROH chez les animaux croisés confirme que certains haplotypes sont partagés entre les animaux Large White et Piétrain. Ceux-ci peuvent se retrouver ensuite à l'état homozygote chez leurs descendants croisés. La présence de ROH identiques entre deux populations porcines de race pure (Large White et Landrace) a précédemment été mise en évidence mais les ROH n'avaient pas été caractérisés chez des éventuels descendants croisés (Zanella *et al*, 2016).

Les haplotypes partagés ici entre les deux races parentales de notre étude sont à l'origine de plusieurs ROH chez leurs descendants croisés dont deux d'entre eux retrouvés chez plus de 50 % de la population de croisés. La première région se situe entre 55969735 et 69928170 pb et la seconde, plus petite, entre 92865347 et 95428483 pb. Concernant les animaux Piétrain, la totalité de la population étudiée ici possède deux ROH sur le chromosome 8, entre 39391675 et 47199839 pb et entre 55969735 et 69661678 pb. La couverture importante en ROH sur le chromosome 8 des Piétrain que nous avons mis en évidence ici, (Figure 5 et 6) avait également été observée lors d'une précédente étude (Bosse *et al*, 2015).



**Figure 6** - Répartition des ROH le long du chromosome 8 chez les trois types génétiques étudiés. Chaque ligne correspond à un individu. Chaque trait plein représente un ROH détecté chez un individu.

LW : Large White, PI : Piétrain ; PLW : Piétrain x Large White  
 Cette visualisation a été réalisée avec la librairie R detectRUNS de Filippo Biscarini.

La distribution et la taille des ROH dans les génomes seraient principalement liées aux phénomènes démographiques intrinsèques aux populations, ainsi qu'aux éléments de recombinaison chromosomique lors de la méiose. Chez le porc, les recombinaisons sont plus fréquentes dans les régions périphériques que dans les régions centrales des chromosomes (Bosse *et al*, 2012). La couverture en longs ROH au niveau de la région centrale du chromosome 8 pourrait donc être liée à la plus faible présence de recombinaisons dans cette région, limitant les cassures possibles.

De plus, l'observation d'haplotypes partagés entre les deux races parentales ou ceux présents chez tous les individus d'une population étudiée, comme ici chez les Piétrain, laissent à penser que ces régions pourraient contenir d'éventuels haplotypes fixés par la sélection. Concernant les ROH présents chez les animaux croisés, nous n'avons pas mis en évidence de gènes majeurs qui auraient pu être fixés dans cette zone du chromosome. Pour les deux ROH partagés entre tous les animaux Piétrain étudiés, le gène KIT (position 41402334 à 41492306 pb) a été identifié dans le premier ROH des Piétrain. KIT est un gène pléiotrope mais connu pour être impliqué dans le déterminisme de la couleur de la robe. Des analyses plus approfondies à la fois de la démographie des populations parentales, des profils de recombinaisons chromosomiques, ainsi que sur les annotations fonctionnelles des gènes pourraient apporter des informations pertinentes sur le maintien et la localisation de ces longs ROH au niveau du chromosome 8.

## CONCLUSION

Cette étude présente la détection de ROH comme outil pour la caractérisation de la diversité génomique dans des populations de race pure en sélection et de leurs descendants croisés. Bien que de l'hétérozygotie soit générée chez les animaux croisés, des régions homozygotes persistent dans leur génome, révélant le partage d'haplotypes entre les populations parentales. Nous avons estimé la consanguinité génomique via le coefficient  $F_{ROH}$ . Celui-ci permet, contrairement au coefficient  $F_X$  estimé à partir des données pedigree, de caractériser plus finement la consanguinité présente chez un individu, à l'échelle du génome mais également à une échelle locale telle que celle du chromosome. Si l'estimation de la consanguinité par pedigree dépend de la qualité et de la profondeur de l'information généalogique, l'estimation de la consanguinité génomique est quant à elle, dépendante de la densité des marqueurs utilisés, de la qualité de génotypage et du contrôle qualité réalisé au préalable. De plus, une attention particulière doit être apportée au choix des paramètres utilisés dans la détection, certains d'entre eux influençant grandement le nombre de ROH détectés. L'analyse de détection de ROH à partir de génotypage devrait donc être complétée par l'analyse des séquences de certains animaux afin de confirmer la présence de ces régions homozygotes.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- Barbato M., Orozco-terWengel P., Tapio M., Bruford M.W., 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.*, 6, 109.
- Bidanel J.-P., 1992. Comment exploiter la variabilité génétique entre races: du croisement simple à la souche synthétique. *Prod. Anim.*, hors série, 249-254.
- Boichard D., 2002. Pedig : a fortran package for pedigree analysis suited for large populations. *World Congr. Genet. Appl. Livest. Prod.*, 7.
- Bosse M., Megens H.-J., Madsen O., Crooijmans R.P.M.A., Ryder O.A., Austerlitz F., Groenen M.A.M., de Cara M.A.R., 2015. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome Res.*, 25, 970-981.
- Bosse M., Megens H.-J., Madsen O., Paudel Y., Frantz L.A.F., Schook L.B., Crooijmans R.P.M.A., Groenen M.A.M., 2012. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. *PLoS Genet.*, 8, e1003100.
- Howard J.T., Tiezzi F., Huang Y., Gray K.A., Maltecca C., 2016. Characterization and management of long runs of homozygosity in parental nucleus lines and their associated crossbred progeny. *Genet. Sel. Evol.*, 48, 91.
- Maignel L., Labroue F., 2001. Analyse de la variabilité génétique des races porcines collectives et des races locales en conservation à partir de l'information généalogique. *Journ. Rech. Porc.*, 33, 111-117.
- Peripolli E., Munari D.P., Silva M.V.G.B., Lima A.L.F., Irgang R., Baldi F., 2017. Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.*, 48, 255-271.
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.*, 81, 559-575.
- VanRaden P.M., 2008. Genomic Measures of Relationship and Inbreeding. *INTERBULL Bull.*, 37, 33-36.
- Zanella R., Peixoto J.O., Cardoso F.F., Cardoso L.L., Biegelmeyer P., Cantão M.E., Otaviano A., Freitas M.S., Caetano A.R., Ledur M.C., 2016. Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. *Genet. Sel. Evol.*, 48, 24.