

Maximisation de la diversité génétique au cours de la création d'une lignée synthétique porcine

Audrey GANTEIL (1,2), Silvia T. RODRIGUEZ-RAMILO (1), Bruno LIGONESCHE (2), Catherine LARZUL (1)

(1) GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan,

(2) SAS NUCLEUS, 35650 Le Rheu

audrey.ganteil@inrae.fr

Maximisation de la diversité génétique au cours de la création d'une lignée synthétique porcine

La création d'une lignée synthétique implique un ou plusieurs croisements entre des races aux aptitudes complémentaires. La représentativité des races parentales chez les animaux croisés ainsi que la génération de diversité génétique sont des paramètres importants à estimer au cours des premières générations de la lignée afin d'assurer sa viabilité à long terme. Aujourd'hui, les données génomiques permettent de caractériser avec précision la structure et la diversité génétique présente dans la population synthétique d'intérêt. L'objectif de cette étude est d'exploiter trois méthodologies décrites dans la littérature afin de maximiser la diversité dans un contexte de création de lignée synthétique porcine issue d'un croisement trois voies. Pour cela, les animaux des deux premières générations de la lignée synthétique ainsi que les fondateurs de race pure ont été génotypés sur une puce SNP de moyenne densité. Nous avons calculé des index d'originalité génomique et estimé l'origine raciale des segments chromosomiques pour chaque animal. Ces données ont été utilisées, en complément de leurs performances, comme outils décisionnels lors du choix des reproducteurs de la seconde génération de la lignée. Les animaux aux index d'originalité les plus élevés et aux origines raciales chromosomiques les plus équilibrées ont été privilégiés. Enfin, pour gérer les accouplements, nous avons estimé un coefficient de parenté génomique basé sur le partage de segments chromosomiques entre deux individus. Ainsi, lors de la création d'une lignée synthétique, l'utilisation des données génomiques pour la sélection des animaux permet, durant les premières générations, de conserver les individus les plus représentatifs de la diversité génétique existante.

Maximizing genetic diversity during the creation of a synthetic porcine line

Creating a synthetic line involves one or more crosses between breeds with complementary abilities. The representativeness of the parental breeds in the crossbred animals as well as the generation of genetic diversity are important parameters to estimate during the first generations in order to ensure long-term viability. Currently, genomic data allow the structure and genetic diversity of a synthetic population to be characterized with precision. The objective of this study was to use three methods described in the literature to maximize diversity in a context of creating a synthetic pig line, using three-way crossbreeding. To this end, animals of the first two generations of the synthetic line as well as the pure-breed founders were genotyped on a medium-density SNP chip. We calculated genomic originality indexes and estimated the racial origin of the chromosomal segments for each animal. These data were used, along with their performance, as a tool to choose the reproducers of the second generation. Animals with the highest originality indexes and the most balanced chromosomal racial origins were favored. Finally, to manage mating, we estimated a genomic kinship coefficient based on the sharing of chromosomal segments between two individuals. Thus, during creation of a synthetic line, using genomic data to select animals during the first generations allows the individuals that most represent the existing genetic diversity to be kept.

INTRODUCTION

La création d'une lignée synthétique (ou composite) peut être envisagée afin de répondre à des besoins actuels ou futurs des filières d'élevage en termes de productivité, de qualité des produits, de robustesse des animaux ou de gestion de leur santé. Il s'agit de croiser deux ou plusieurs races aux aptitudes généralement complémentaires (Bidanel, 1992). Après deux ou trois générations sans sélection, la lignée est conduite en population fermée puis des objectifs de sélection spécifiques sont définis (Legault *et al.*, 1996).

Lors de la création d'une nouvelle lignée composite, il est pertinent d'évaluer la variabilité génétique générée suite au croisement et de gérer la représentativité des génomes des races fondatrices au fil des générations. Ces problématiques s'intègrent dans l'objectif de produire une lignée durable s'appuyant sur un progrès génétique à long terme. L'un des points clé est donc de maximiser la diversité génétique présente dans les premières générations du croisement. D'un point de vue stratégique, peu d'éléments sont disponibles dans la littérature pour répondre à cet objectif dans un contexte de création de lignée synthétique.

Ce travail est lié au projet de création d'une nouvelle lignée paternelle par l'entreprise de sélection NUCLEUS. Cette lignée est issue d'un croisement trois voies et s'oriente vers des objectifs d'amélioration de la croissance, de la qualité de la viande et de la vigueur du porcelet. De plus, dans ce projet, une attention particulière est portée à la gestion de la diversité génétique, grâce notamment à l'accessibilité du génotypage des populations porcines.

L'objectif de cet article est de présenter les outils d'analyse génomique mis en place au cours de la création de la lignée afin de conserver une représentativité maximale de la diversité génétique lors des premières générations.

1. MATERIEL ET METHODES

1.1. Dispositif de création de la lignée synthétique

Le dispositif de croisement a été mis en place au sein d'élevages de l'entreprise de sélection NUCLEUS. Pour la première étape, des femelles Large White ont été inséminées avec des verrats Piétrain. Les truies croisées Piétrain x Large White (PLW) ont ensuite été accouplées avec des verrats Duroc. Les animaux issus de ce croisement ont constitué la première génération, G0, de la lignée synthétique. Pour les générations suivantes, la lignée est conduite en population fermée (Figure 1).

Afin de répondre à l'objectif de maximisation de la diversité génétique, les accouplements des animaux de la génération G0 ont été programmés en tenant compte de la généalogie pour minimiser l'élévation de consanguinité chez les descendants. Nous avons utilisé pour cela la méthode des contributions optimales (Meuwissen, 1997).

A la génération suivante (G1), la composition génomique est en moyenne un quart Piétrain et un quart Large White et à moitié Duroc. Cependant, l'aléa de méiose ainsi que les recombinaisons chromosomiques sont à l'origine de compositions génomiques variées autour de cette moyenne. Les données généalogiques étant peu informatives à ce stade, nous avons choisi d'exploiter des données génomiques afin d'analyser avec précision non seulement la diversité présente dans la population G1 mais aussi la composition individuelle des G1 en haplotypes des trois races fondatrices.

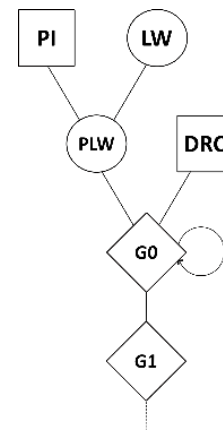


Figure 1 – Schéma de constitution de la lignée synthétique.

DRC: Duroc, G0: (Piétrain x Large White) x Duroc, G1: G0 x G0, LW: Large White, PI: Piétrain, PLW: Piétrain x Large White

1.2. Génotypes et contrôle qualité des données

Dans le cadre de la création de la lignée synthétique, l'entreprise de sélection NUCLEUS a collecté des biopsies d'oreille des animaux de race pure et croisés. A partir de ces biopsies, l'ADN des animaux a été extrait puis les génotypages ont été réalisés sur les puces porcines Illumina SNP60 versions Geneseek et Choice. Un contrôle qualité a ensuite été réalisé avec le logiciel PLINK version 1.9 (Purcell *et al.*, 2007). Seuls les marqueurs présents sur les autosomes ont été conservés. Nous avons contrôlé 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, 48621 marqueurs et 5073 animaux ont été conservés pour l'analyse : 81 Piétrain, 243 Large White, 89 Duroc, 443 PLW, 758 G0 et 3459 G1.

Dans notre protocole, l'ensemble des animaux G1 a été génotypé. L'objectif était de disposer des informations génomiques au même moment que le contrôle des performances des animaux. Les données génomiques ainsi que les performances étaient ensuite prises en compte pour choisir les futurs reproducteurs. Ainsi, les analyses définies dans les paragraphes suivants sont réalisées chaque mois afin de disposer des informations génomiques avant chaque contrôle de performances. Les résultats présentés ici sont issus des analyses réalisées en Juin 2020.

1.3. Distances génétiques entre les individus

Un positionnement multidimensionnel (MDS) de l'ensemble de des animaux de la population étudiée a été réalisé avec le logiciel PLINK 1.9. Cette analyse permet d'estimer les distances génomiques entre les individus. Pour chaque paire d'animaux, les distances sont exprimées en comptage de marqueurs différents entre les deux animaux. Cette analyse a ensuite été représentée dans un plan à deux dimensions.

1.4. Origine raciale des fragments chromosomiques

Afin d'estimer chez chaque individu G1 la fraction allélique issue de chacune des trois races fondatrices de la lignée, nous avons appliqué la méthode BOA (Breed Origin to Allele) (Vandenplas *et al.*, 2016). Cette méthode consiste dans un premier temps à déterminer les haplotypes spécifiques de chacune des races pures étudiées puis d'identifier ces haplotypes chez les individus croisés issus de ces races pures. Le programme

informatique BOA permet d'assigner à une race pure les allèles d'individus croisés seulement de la première génération d'un croisement trois voies. Or, dans notre cas, pour la caractérisation des G1 (seconde génération), nous avons dû faire le choix de considérer les animaux PLW comme une origine unique. Cela nous permet d'obtenir en G1 les pourcentages d'assignation PLW et Duroc des allèles.

1.5. Index d'originalité

Nous avons calculé pour chaque animal G1 un index d'originalité ou ORI (Danchin-Burge *et al.*, 2016). L'ORI permet de comparer le génotype de chaque animal par rapport aux fréquences alléliques moyennes de la population de référence. Le principe est de mettre en évidence des individus porteurs d'allèles les moins fréquents dans la population d'intérêt. Par exemple, si l'on considère un animal i , au génotype AB et CC pour deux locus bialléliques dont les fréquences alléliques dans la population d'intérêt sont f_A , f_B et f_C :

$$f_i = (f_A + f_B + f_C + f_C)/4$$

Ici, f_i , représente la fréquence moyenne des marqueurs portés par l'individu i . f_i , est plus élevé chez les individus aux allèles les plus fréquents. Ensuite, on estime f_{pop} , moyenne des fréquences moyennes des marqueurs pour l'ensemble des n individus de la population :

$$f_{pop} = \sum_{i=1}^n f_i/n$$

On peut ensuite calculer l'index ORI :

$$ORI_i = -\frac{f_i - f_{pop}}{\sigma_{f_{pop}}}$$

Où $\sigma_{f_{pop}}$ est l'écart type des moyennes des fréquences alléliques aux marqueurs dans la population. Ainsi, un animal avec un ORI positif sera considéré comme original, il aura tendance à porter des allèles plutôt peu fréquents dans la population. Inversement un animal à l'ORI négatif sera porteur d'allèles plutôt communs dans la population.

Ces index ORI sont calculés après obtention des fréquences alléliques des SNP dans la population d'individus G1 avec le logiciel PLINK 1.9.

La corrélation entre l'index ORI et le pourcentage d'allèles issus du Duroc chez les animaux G1 a été réalisée avec le logiciel R (R Core Team, 2018).

1.6. Gestion des accouplements

Après le choix des futurs reproducteurs mâles et femelles nous avons défini un protocole afin de gérer leurs accouplements. Celui-ci se base sur la détection de segments chromosomiques partagés chez de futurs couples de mâles et de femelles. Plus un couple partagera des segments chromosomiques et plus il existera un risque que ces derniers soient générateurs de

segments chromosomiques à l'état homozygote chez les descendants. Il s'agit ici de la notion de « Runs of Homozygosity » ou ROH que nous avons abordé dans une étude précédente chez les fondateurs Piétrain, Large White et PLW de la lignée (Ganteil *et al.*, 2020). Ainsi, pour le choix des accouplements, nous calculons un coefficient de parenté génomique basé sur l'estimation des segments chromosomiques partagés entre les futurs parents, f_{SEG} (de Cara *et al.*, 2013) :

$$f_{SEG_{ij}} = \sum_k \sum_{a_i=1}^2 \sum_{b_j=1}^2 L_{IBD_k}(a_i, b_j)/4L$$

Où $f_{SEG_{ij}}$ est le coefficient de parenté génomique entre les individus i et j , $L_{IBD_k}(a_i, b_j)$ est la taille du k -ième segment partagé mesuré entre le chromosome homologue a de l'individu i et le chromosome homologue b de l'individu j . L'estimation des segments partagés k est réalisée avec le logiciel Germline (Gusev *et al.*, 2008) après le phasage des génotypes avec Fimpute (Sargolzaei *et al.*, 2014).

2. RESULTATS ET DISCUSSION

2.1. Structure génomique de la population

Afin d'observer les distances génomiques entre les individus, nous avons réalisé un MDS plot avec les animaux fondateurs de la lignée ainsi que les animaux des deux premières générations, G0 et G1 (Figure 2). Cette représentation en deux dimensions nous permet de visualiser les différents types génétiques en cinq sous-groupes distincts. L'axe 1 différencie les animaux Duroc des deux autres populations de fondateurs Large White et Piétrain. Ces deux dernières populations sont discriminées par l'axe 2. A mi-distance entre les populations Large White et Piétrain se situent les individus PLW, illustrant la composition chromosomique à moitié Piétrain et à moitié Large White de ces animaux.

Deux individus, l'un PLW et le second Large White, sont distants par rapport au groupe auquel ils devraient être associés. Cette observation est expliquée par une erreur de filiation de ces animaux que nous avons mis en évidence au cours de l'étude. En effet, l'animal Large White est en réalité trois-quarts Large White et un quart Piétrain. Quant à l'animal PLW, il s'agit d'un animal trois-quarts Piétrain et un quart Large White. Dans notre projet, ce type de graphique est donc très utile afin de repérer rapidement certaines erreurs de filiation lorsque les parents sont issus de deux types génétiques différents.

Les animaux G0 sont situés entre leurs deux populations parentales, Duroc et PLW. Les animaux G1 possèdent quant à eux, une structure de population beaucoup plus dispersée que leurs parents G0 traduisant les phénomènes d'aléa de méiose et de recombinaisons chromosomiques à l'origine de la diversité au sein de cette génération.

2.2. Choix des futurs reproducteurs

Au moment du contrôle de performances, nous analysons à la fois les pourcentages d'assignation PLW et Duroc des allèles et les index ORI. Ces données sont étudiées en parallèle des performances des animaux afin d'éviter de conserver des animaux aux performances trop en dessous de la moyenne de la population.

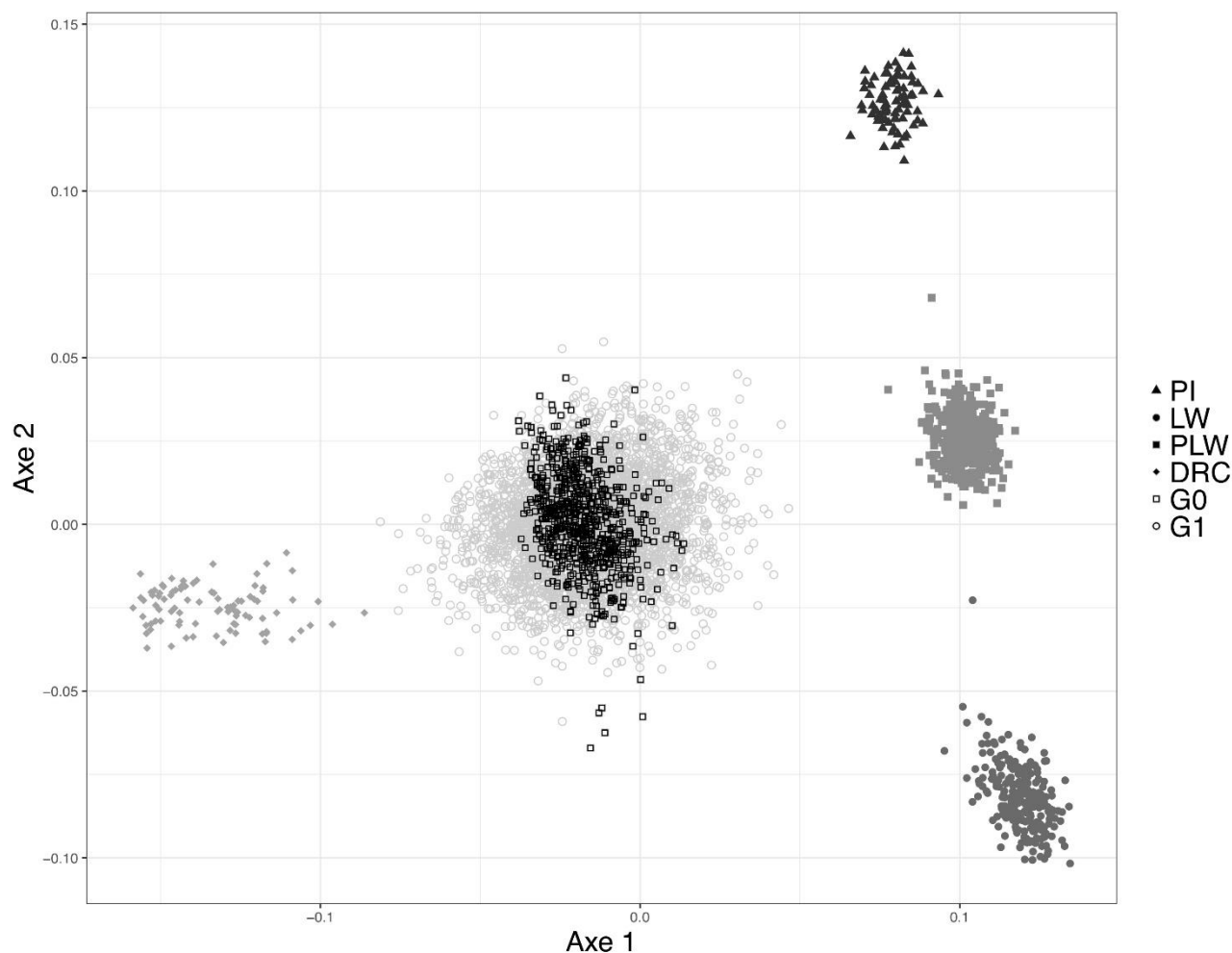


Figure 2 – Distances génétiques entre l'ensemble des individus du dispositif de création de la lignée synthétique, fondateurs et animaux croisés.

L'analyse d'assignation des allèles, nous permet d'obtenir par animal G1 les pourcentages d'assignation des allèles aux fondateurs Duroc ou PLW. Le Tableau 1 résume les principales statistiques concernant les fréquences d'assignation totale ainsi que l'assignation PLW et Duroc. La méthode employée ici nous permet d'assigner en moyenne 85 % des allèles. Les allèles qui ne seraient pas assignés peuvent correspondre à des haplotypes partagés entre les races qu'il est difficile de discriminer entre les Duroc et PLW. Concernant les assignations des allèles en Duroc ou PLW, on obtient des résultats similaires pour les deux types fondateurs. En moyenne, nous assignons un nombre de SNP similaires à l'origine Duroc et à l'origine PLW (43 %). Cela est cohérent avec le pourcentage d'assignation totale moyen et la composition génomique des parents qui sont à moitié PLW et à moitié Duroc.

Tableau 1 – Statistiques descriptives sur l'assignation raciale des allèles chez les animaux G1.

	Minimale	Maximale	Moyenne	Ecart-type
Assignation totale	0,69	0,92	0,85	0,027
Assignation PLW	0,20	0,66	0,43	0,057
Assignation DRC	0,23	0,62	0,43	0,054

Pour le choix des futurs reproducteurs, nous essayons de conserver les animaux qui ont des pourcentages d'assignation

en Duroc proches de 50 %. De plus, la structure familiale est prise en compte pour les mâles en conservant un fils par père. Concernant les femelles, un effectif minimum de femelles à conserver est fixé. Un minimum de sélection phénotypique est mis en place ici car seuls les animaux aux performances supérieures à un certain seuil pour des caractères de croissance et de composition corporelles sont conservés.

L'accent mis sur le Duroc est ici un choix de l'entreprise de sélection. De plus, les index ORI sont pris en compte, nous essayons de favoriser les animaux aux index positifs, soit les animaux présentant plutôt des allèles moins fréquents dans la population. D'un point de vue théorique, en appliquant indéfiniment cet index, on devrait finir par avoir des fréquences d'un demi pour tous les allèles, ce qui constituerait la diversité génétique maximale que l'on peut atteindre. Cependant, le coefficient de corrélation entre l'ORI et le pourcentage d'assignation en Duroc est de -0,44 (Figure 3). Cette corrélation négative peut s'expliquer par la composition chromosomique des animaux G0 et G1. En effet, ces animaux sont à moitié Duroc et en moyenne un quart Piétrain et un quart Large White. Il y a donc plus de probabilité pour un animal G1 d'avoir hérité d'un allèle spécifique au Duroc que d'un allèle spécifique du Piétrain ou du Large White. Par conséquent les allèles d'origine Duroc seront prépondérants dans la population G1. Les animaux porteurs de ces allèles Duroc seront le plus souvent associés à des ORI négatifs. Il est donc nécessaire de considérer ces deux paramètres génomiques conjointement.

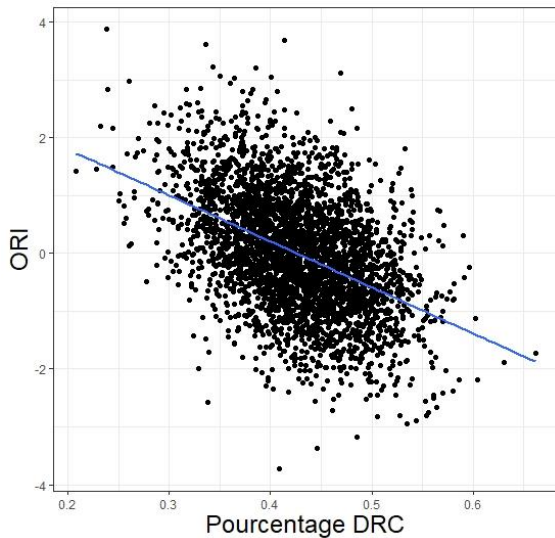


Figure 3 – Corrélations entre l’ORI et le pourcentage d’assignation des allèles en Duroc pour les animaux G1.

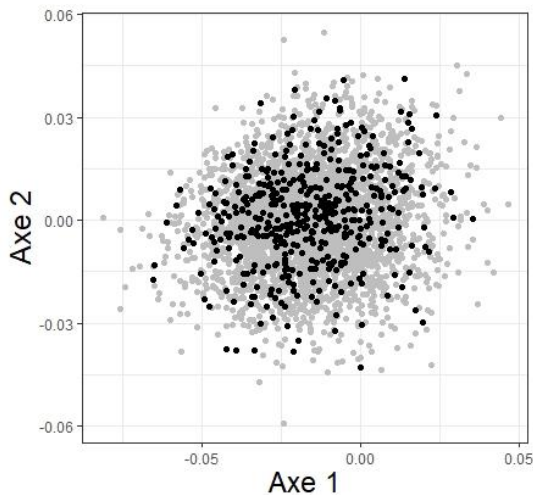


Figure 4 – Distances génétiques entre l’ensemble des individus G1 qui ont été génotypés. Les individus conservés en tant que futurs reproducteurs sont en noir.

Afin de suivre l’évolution de la structure de la population G1 avant et après le choix des futurs reproducteurs, nous avons réalisé un MDS plot afin de visualiser les distances génomiques entre les animaux (Figure 4). Ce graphique nous permet de distinguer parmi l’ensemble des G1 génotypés ceux qui sont conservés pour devenir de futurs reproducteurs. Ces derniers se répartissent globalement de manière homogène au sein du nuage de points illustrant les animaux G1 génotypés. Ce résultat est encourageant pour la suite de notre projet. En effet, il semblerait que l’exploitation des données génomiques lors du tri des animaux nous permette déjà de conserver des individus représentatifs de la population G1.

2.3. Choix des accouplements

Bien que le tri soit une étape cruciale afin de parvenir à notre objectif de maximisation de la diversité au cours des premières générations de la nouvelle lignée, une seconde étape doit également être prise en considération : la gestion des accouplements. Pour cette dernière, nous avons choisi une stratégie de minimisation de la consanguinité génomique à la génération suivante. Le coefficient de parenté que nous

calculons (f_{SEG}) est basé sur le partage de segments chromosomiques entre deux animaux. Puis nous sélectionnons les couples minimisant ce coefficient. L’utilisation de ce coefficient de parenté nous semble pertinente puisqu’il permet de limiter l’apparition de segments homozygotes chez les descendants qui pourraient être à l’origine d’une diminution de la diversité.

Pour la génération suivante, la G2, nous pensons poursuivre cette même stratégie de gestion à la fois pour le choix des reproducteurs et la gestion de leurs accouplements. L’analyse des génotypages des animaux G2 nous permettra également de faire un bilan sur la diversité génétique existante à cette génération. Nous envisageons par exemple d’observer la distance génétique entre les animaux mais également de quantifier la proportion de segments homozygotes présente à cette génération.

Pour finaliser la création de la lignée, il va être indispensable de s’interroger sur le nombre de générations durant lequel nous allons appliquer ce protocole de maximisation de la diversité. Il est généralement admis qu’au bout de trois ou quatre générations, le brassage est suffisant pour considérer la population comme homogène et débiter la sélection. Au niveau génomique, il n’existe pas à l’heure actuelle, de critère permettant de quantifier l’homogénéité d’une population synthétique. Il convient donc maintenant de travailler sur cette problématique afin d’estimer des paramètres informatifs de cette notion d’homogénéité génomique. Nous envisageons de suivre l’évolution de la taille des haplotypes spécifiques des races parentales ainsi que d’étudier les événements de recombinaison.

CONCLUSION

Dans cet article, nous proposons des méthodes d’analyse de données génomiques afin de maximiser la diversité génétique au cours d’un protocole de création d’une lignée synthétique. A notre connaissance, il s’agit de la première étude en porc sur ce sujet. Nos résultats soulignent que les données génomiques permettent d’estimer la diversité génomique avec précision. Nous pouvons obtenir une estimation chez chaque individu du pourcentage de segments chromosomiques issus des fondateurs Duroc et PLW ainsi qu’une indication sur la présence chez un individu d’allèles peu fréquents dans la population. Ces outils ont été mis en application pour la gestion de la diversité au cours de la génération G1 lors du tri des animaux. Ils semblent être efficaces afin de conserver des futurs reproducteurs représentatifs de la diversité génétique présente en G1. Nous envisageons de poursuivre une stratégie similaire de maximisation de la diversité à la génération suivante la G2. La prochaine étape de ce projet sera le développement d’un programme de sélection spécifique à la lignée.

REMERCIEMENTS

Nous tenons à remercier Aurélie Le Dréau et Emilie Normand du service R&D de NUCLEUS pour la gestion des biopsies nécessaire à ce projet. Merci également à Vincent Cousin et Laurent Guéry du service technique de NUCLEUS responsables du tri et de l’organisation des accouplements de la nouvelle lignée ainsi qu’à tous les éleveurs impliqués dans ce projet. Nous remercions aussi Jérémie Vandenplas (Wageningen University & Research, Animal Breeding & Genomics) pour le partage des exécutables afin d’utiliser la méthode BOA.

REFERENCES BIBLIOGRAPHIQUES

- Bidanel J.-P., 1992. Comment exploiter la variabilité génétique entre races: du croisement simple à la souche synthétique. INRA Prod. Anim., hors série "Eléments de génétique quantitative et application aux populations animales", 249-254.
- Danchin-Burge C., Moureaux S., Boichard D., Baur A., Fritz S., 2016. Un indicateur innovant de variabilité génétique à partir des données moléculaires. Conférence "Congrès Mondial de la race Brune", Mende, France.
- Gusev A. et al, 2008. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.*, 19, 318-326.
- Legault C., Ménéssier F., Mérat P., Ricordeau G., Rouvier R., 1996. Les lignées originales de l'INRA: Historique, développement et impact sur les productions animales. INRA Prod. Anim., hors série, 41-56.
- Meuwissen T.H., 1997. Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci.*, 75, 934.
- Purcell S. et al, 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.*, 81, 559-575.
- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, Consultable : <https://www.R-project.org>.
- Sargolzaei M., Chesnais J.P., Schenkel F.S., 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15, 478.
- Vandenplas J., Calus M.P., Sevillano C.A., Windig J.J., Bastiaansen J.W., 2016. Assigning breed origin to alleles in crossbred animals. *Genet. Sel. Evol.*, 48, 61.