Potential of genomic and intestinal microbiota information for the selection of feed efficiency in pigs

Amir Aliakbari

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National Polytechnique de Toulouse

Présentée et soutenue par

Amir ALIAKBARI

Le 5 mai 2021

Potentiel des informations de la génomique et du microbiote intestinal pour la sélection de l'efficacité alimentaire en porc

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Pathologie, Toxicologie, Génétique et Nutrition**

Unité de recherche :
GenPhySE- Unité Génétique, Physiologie et Systèmes d'Elevage

Thèse dirigée par
Hélène GILBERT

Jury

Mme Torres Sweeney, Rapporteure
M. Jörn Bennewitz, Rapporteur
Mme Noelia Ibañez-Escriche, Examinatrice
Mme Claire Rogel-Gaillard, Examinatrice
Mme Hélène GILBERT, Directrice de thèse

RESUME

L'objectif principal de la thèse était d'étudier comment les informations génomiques de l'animal et de son microbiote peuvent contribuer à améliorer la sélection pour l'efficacité alimentaire chez le porc. La thèse s'est appuyée sur les données de deux lignées de porcs issues de 10 générations de sélection divergente pour l'efficacité alimentaire. En plus de phénotypes enregistrés pour environ 200 porcs par lignée et par génération, 588 échantillons de fèces ont été collectés en générations 9 et 10. De plus, des génotypes pour environ 1000 animaux par lignée étaient disponibles. Cinq caractères ont été étudiés : la consommation moyenne journalière résiduelle, l'indice de consommation, la consommation moyenne journalière, le gain moyen quotidien et l'épaisseur de lard dorsal. Dans cette thèse, nous avons montré que les informations moléculaires sur les porcs ou leur microbiote peuvent améliorer la sélection pour l'efficacité alimentaire. Ce caractère est coûteux à enregistrer, alors que les informations moléculaires pourraient être plus faciles à obtenir sur un grand nombre de porcs. Dans ce projet, le potentiel du génotypage des animaux a été examiné dans le premier chapitre, et celui du microbiote intestinal a été exploré dans les deux suivants. Nous avons d'abord montré que lorsque la disponibilité des données est limitée, la prédiction génomique avec une population de référence combinant des animaux de lignées génétiquement liées peut être aussi précise qu'une prédiction génomique utilisant une population de référence de la lignée cible uniquement. Comparant de nombreux scénarios, nos résultats ont fourni des repères pour la construction de populations de référence pour initier la sélection génomique dans des lignées petites, qui ne disposent pas d'un grand nombre d'échantillons ou de données historiques et sont développées simultanément. Cette situation peut être rencontrée en volaille et en porc ainsi que dans d'autres populations en croisement. Des études complémentaires seront nécessaires pour quantifier le potentiel économique de cette approche et clarifier l'équilibre optimal entre génotypage et de phénotypage. Dans les chapitres suivants, nous avons montré que la variabilité du microbiote intestinal, captée par séquençage partiel du gène de l'ARNr 16S, contribue à la variabilité des caractères de production, en particulier de l'efficacité alimentaire. Dans un premier temps, nous avons identifié des composantes du microbiote (genres, OTU, indices de α-diversité) héritables (48 genres sur les 75 analysés, plus deux indices de α-diversité). Vingt et un

de ces genres, appartenant aux familles *Lachnospiraceae*, *Ruminococcaceae*, *Prevotellaceae*, *Lactobacillaceae*, *Streptococcaceae*, *Rikenellaceae* et *Desulfovibrionaceae*, et les deux indices de α-diversité étaient génétiquement corrélés à certains caractères. Deuxièmement, l'étude de la microbiabilité a montré une contribution substantielle des effets microbiote à la variabilité de l'efficacité alimentaire (> 10%) et une contribution négligeable pour les autres caractères (< 5%). De plus, cette étude a révélé que la génétique de l'hôte avait une contribution plus élevée que le microbiote à la variance des caractères étudiés (héritabilité plus élevée que les valeurs de microbiabilité). Cette dernière étude a également montré des associations significatives de certains taxons microbiens avec les performances. Ces résultats ont souligné la possibilité d'utiliser certains caractères microbiens comme marqueurs pour la sélection de l'efficacité alimentaire chez les porcs. Des études complémentaires seront nécessaires pour évaluer comment les informations génomiques de l'hôte et du microbiote peuvent être combinées dans des modèles de prédiction pour soit mieux prédire les valeurs génétiques elles-mêmes, soit même obtenir des prédictions conjointes des valeurs génétiques et microbiote, qui conduiraient à la sélection de l'hologénome pour une efficacité de production améliorée.

ABSTRACT

The main objective of the thesis was to investigate how genomic tools applied to the animal and its microbiota can contribute to improving selection for feed efficiency in pigs. The thesis relied on data from two pig lines from 10 generations of divergent selection for feed efficiency. Together with phenotypic records for about 200 pigs per line in all generations, 588 feces samples from generations 9 and 10 were collected. In addition, SNP genotyping data for about 1000 animals per line were available. Five traits were investigated: residual feed intake, feed conversion ratio, daily feed intake, average daily gain and backfat thickness. Throughout the thesis, we showed that molecular information acquired on the pigs or their microbiota could improve selection for feed efficiency. This trait is costly to record, whereas molecular information could be easier to obtain on a large number of pigs. In this project, the potential of genomic tools applied to pigs was examined in the second chapter, and it was explored in the two subsequent ones for the gut microbiota. We then first showed that when data availability is limited, genomic prediction using a training set combining animals from genetically related lines can be as accurate as genomic prediction using a training set from the target population only. Based on numerous scenario comparisons, our results provided insights into the design of reference populations to initiate genomic selection in livestock lines with small population size, do not have a large number of historical samples or data, and are developed simultaneously, as can be encountered in poultry and pig breeding, as well as in other crossbreeding schemes. Further studies would be needed to assess the economic potential of this approach and clarify the optimum balance between genotyping and phenotyping efforts. In the following chapters, we showed that the gut microbiota variability, captured via partial 16S rRNA gene sequencing, contributes to the variability of production traits, in particular feed efficiency traits. First, we identified microbiota components (genera, OTU, α-diversity indexes) with significant heritability (48 genera out of the 75 analysed, plus two α-diversity indexes). Twenty-one of these genera, belonging to the *Lachnospiraceae*, *Ruminococcaceae*, *Prevotellaceae*, *Lactobacillaceae*, *Streptococcaceae*, *Rikenellaceae* and *Desulfovibrionaceae* families, and the two α-diversity indexes were genetically correlated with some of the traits. Second, the study of the microbiability showed a substantial contribution of the microbial effects on the

variability of feed efficiency traits (> 10%) and negligible contribution for other traits (<5%). In addition, this study revealed that host genetics had a higher contribution than the microbial community to the variance of the studied traits (higher heritability than microbiability values). This last study also showed significant associations of some microbial taxa with feed efficiency and performance traits. These results pointed out the possibility of using some microbial traits as markers for the selection of feed efficiency in pigs. Further studies will be needed to evaluate how genomic information of the host and the microbiota can finally be combined in prediction models to either better predict the breeding values themselves, or even obtain joint predictions of breeding and microbiota values, that would lead to the selection of the hologenome for improved production efficiency.

Acknowledgements

Spending more than three years in the GenPhySE lab at INRAE gave me the chance of meeting many nice people. With all difficulties and happy moments, it was a wonderful experience for me to see the different faces of life.

I would like first, and foremost to thank my supervisor Hélène, who provided the encouragement and advice that were necessary to proceed through my PhD program and to complete my dissertation. Thank you so much for all the advice, ideas, moral supports, and patience while guiding me throughout this way. Thank you for all the efficient weekly meetings that we had, which always started with a warm "How are you" from you. I learn a lot from you. Your comments on my manuscripts were always very helpful. You always kept me motivated during my PhD journey. Thanks for your optimism, hard work, and for being always there, during the entire thesis and for sure during the writing, without your support it would have been an impossible task. Also, thank you for helping me with translating all French documents and filling the forms and my job applications. Thank you for the valuable advice you gave me about job applications. Finally, a huge thank you to teach me how to be a real researcher.

Special thanks to Juliette Riquet who was always very friendly with me and gave me positive energy with her nice talks and smiles.

Special thanks to Ingrid David that involved me in her brilliant study. Our talks during the break times always gave me confidence. You are an effective person and thanks for being always so nice.

Andres and Zulma, thank you so much for remembering and inviting me for the wonderful dinners, especially the Christmas dinners. Both of you were extremely kind to me.

Huge thanks to Roberta Rostellato that always helped me in everything. You made my life much easier here.

My office mate Audrey Ganteil, thank you very much for listening to me, for your help, and for being an extremely understanding person.

I would like to especially thank all members of the MicroFeed team and all my friends here in GenPhySE for their help, supports, chats, enthusiasm, and friendship.

In the end, my dear and kind parents, words are very limited to thank you. Thank you very much for all your hard works. I am proud of you. I owe all my success to your sacrifices and supports. Dear mother, thank you so much for never forgetting me, even for a while. Dear father, thank you so much for teaching me to believe that I have the ability to achieve my lofty goals and to always be resilient in the most difficult conditions of life. Special thanks also to my kind sister, Anis, who always made me smile with her funny talks and informed me about the last news of the family (my BBC!).

Amir Aliakbari

12/March/2021

Auzeville-Tolosane

Table of contents

1. General introduction 1
1.1. Importance of feed in the pig industry 2
1.1.1. Feeding costs in the pig industry 3
1.1.2. Environmental impacts of nutrient excretion 4
1.2. Definitions of feed efficiency and indicator traits 4
1.3. Biological basis of feed efficiency 6
1.4. Means to improve the feed efficiency of pigs 8
1.4.1. Nutritional strategies 8
1.4.2. Genetic improvement of feed efficiency 10
1.4.2.1. Aspects of selection for feed efficiency 10
1.4.2.2. Genomic selection in pig breeding programs 14
1.4.3. Improving feed efficiency through digestion efficiency 18
1.4.3.1. Influence of gut microbiome on feed efficiency 18
1.4.3.2. Effect of host genetic on gut microbiome composition 21
1.5. Objectives 23
1.6. References 25
2. Potential of the genomic information to improve selection for feed efficiency 39
2.1. Introduction 40
2.2. Article I: The impact of training on data from genetically-related lines on the accuracy of genomic predictions for feed efficiency traits in pigs 42
2.2.1. Abstract 43
2.2.1.1. Background 43
2.2.1.2. Results 43
2.2.1.3. Conclusions 43

2.2.2. Background 44
2.2.3. Methods 46
2.2.3.1. Population and data structure 46
2.2.3.2. Combining and standardizing traits 49
2.2.3.3. Single nucleotide polymorphism (SNP) genotyping data and imputation 50
2.2.3.4. Model and analyses 50
2.2.3.5. Scenarios 52
2.2.3.6. Accuracy and bias of genomic predictions 54
2.2.3.7. Relationships between training and validation sets 55
2.2.4. Results 55
2.2.4.1. Variance components 55
2.2.4.2. Prediction accuracies 56
2.2.4.3. Prediction biases 58
2.2.4.4. Relationships between and within training and validation sets 58
2.2.5. Discussion 61
2.2.5.1. Computation of prediction accuracies and biases 61
2.2.5.2. Prediction accuracies for production traits 62
2.2.5.3. Prediction accuracies depending on the training set composition 63
2.2.5.4. Impact of selection on accuracy and bias of predictions 64
2.2.5.5. Bias of genomic predictions 65
2.2.6. Conclusions 67
2.2.7. References 68
2.2.8. Appendix 2.1 74
2.2.9. Appendix 2.2 76
2.2.10. Appendix 2.3 77

3. Genetic basis of the gut microbiota and their relationships with production trait 79
3.1. Introduction 80
3.2. Article II: Genetic relationships between feed efficiency and gut microbiome in pig lines selected for residual feed intake 81
3.2.1. Abstract 82
3.2.2. Introduction 83
3.2.3. Materials and Methods 84
3.2.3.1. Data structure 84
3.2.3.2. Faeces sampling, microbial DNA extraction, 16S rRNA gene sequencing and sequence pre-processing 86
3.2.3.3. Statistical analyses 88
3.2.4. Results 91
3.2.4.1. Gut microbiome differences between lines 91
3.2.4.2. Heritability estimates of microbiota traits 95
3.2.4.3. Genetic correlations of microbiota traits with production traits 96
3.2.5. Discussion 102
3.2.5.1. Some genera are under genetic control 102
3.2.5.2. Some genera are genetically correlated with production and FE traits 105
3.2.5.3. α-diversity indexes are under genetic control and are related to FE traits 107
3.2.5.4. Potential for selection and management in pig production 108
3.2.6. Conclusion 108
3.2.7. References 110
3.2.8. Appendix 3.1 117
3.2.9. Appendix 3.2 121
4. How microbiota contributes to the variability of the production traits 129

4.1. Introduction 130
4.2. Article III: Microbiability and microbiome wide association studies for feed efficiency and performance traits in pigs 131
4.2.1. Abstract 132
4.2.2. Introduction 133
4.2.3. Materials and Methods 134
4.2.3.1. Population structure, studied traits and sampling 134
4.2.3.2. Microbial information 136
4.2.3.3. Statistical analyses 137
4.2.3.3.1. Estimation of variance components 137
4.2.3.3.2. Microbiome wide association studies 139
4.2.3.3.3. Significance threshold for MWAS 140
4.2.4. Results 141
4.2.4.1. Estimation of variance components 141
4.2.4.2. Microbiome wide association studies (MWAS) 143
4.2.5. Discussion 147
4.2.5.1. Estimation of variance components 147
.4.2.5.2 Microbiome wide association results 148
4.2.6. Conclusion 150
4.2.7. References 151
4.2.8. Appendix 4.1 155
5. General discussion 157
5.1. Overview of the thesis 158
5.2. Efficiency of genomic evaluation using multi-population training sets 159
5.3. Dependency of feed efficiency variation to the intestinal microbial composition 161

5.4. Selection of host animals based on microbiome evaluations 164
5.5. Conclusion 167
5.6. References 168

List of figures

Figure 1-1. Meat production by type of livestock, World, 1961 to 2018 2

Figure 1-2. Cost of production (€/kg hot carcass weight), split into cost categories in selected EU countries on a closed cycle pig farm 3

Figure 1-3. Dietary energy sources and energy use in the pig 7

Figure 2-1. Design of scenarios to predict validation animals in HRFI (a) and LRFI (b) lines 52

Figure 2-2. Correlations between GEBVp and GEBVw, and their SE as error bars for the HRFI (a) and LRFI (b) lines. 57

Figure 2-3. Bias (regression coefficients of GEBVw on GEBVp) for the HRFI (**a**) and LRFI (**b**) lines. 59

Figure 2-4. Average, minimum and maximum relationship coefficients in the H matrix between individuals of the validation set, and individuals of the training set from the target line and from the reverse line, for **a** and **b** the HRFI target line, for **c** and **d** the LRFI target line 60

Figure 3-1. Abundance percentage of the 75 genera in the LRFI and HRFI lines. 91

Figure 3-2. Non-metric multidimensional scaling (NMDS) based on the Bray-Curtis dissimilarity matrix of the genera community (a) and box plots of the individual coordinates per line on the two first axes of the NMDS (LRFI=low residual feed intake; HRFI= high residual feed intake), with P-value of the ANOVA test of the line differences (b) 93

Figure 3-3. Projection of the genera on the first and second dimensions in a non-metric multidimensional scaling (NMDS) applied to the Bray-Curtis matrix of the genera abundances. The arrows are colored based on the contribution of each genus to the plan. 94

Figure 3-4. Box plots of Shannon and Simpson α-diversity indexes per line 95

Figure 3-5. Box plots of genera abundances per line (LRFI, low residual feed intake; HRFI, high residual feed intake) and *p*-value of ANOVA test of the line differences 101

Figure 4-1. Results of microbiome wide association between operational taxonomic units and residual feed intake (a) and feed conversion ratio (b) 144

Figure 4-2. Results of microbiome wide association study between operational taxonomic units and daily feed intake (a), average daily gain (b) and back fat thickness (c)..... 145

Figure 5-1. Distribution of feed-derived energy between the host and gut bacteria 162

List of tables

Table 1-1. Heritability (h^2) and genetic correlations (r_G) of FCR and RFI in different pig breeds 13

Table 1-2. Genomic prediction accuracy of growth and feed efficiency traits in some studies on pigs 17

Table 2-1. Numbers of animals in the pedigree and data structure 48

Table 2-2. Descriptive statistics of the data for the studied traits in the HRFI and LRFI lines 49

Table 2-3. Number of genotyped animals in the training and validation sets for the six scenarios for the HRFI and LRFI validation sets 53

Table 2-4. Estimates of variance components (SE) of the studied traits 56

Table 3-1. Number (N), minimum (Min), maximum (Max), mean and standard deviation (SD) of the studied traits† in the low residual feed intake (LRFI) and high RFI (HRFI) lines 86

Table 3-2. Estimated heritability (h^2) and standard errors (SE) and descriptive statistics (minimum (Min), maximum (Max), mean and standard deviation (SD)) of α-diversity indexes and genera abundances 96

Table 3-3. Genetic correlations† (SE) of α-diversity indexes and genera with production traits‡ 100

Table 4-1. Descriptive statistics of data structure for the studied traits 136

Table 4-2. Posterior means (± posterior standard deviation) of variance components, heritability and microbiability values of production traits using the four models, and corresponding deviance information criterion (DIC) of each model 142

Table 4-3. Posterior means (± posterior standard deviation) of the main parameters obtained from bivariate analyses between traits with model 4 143

Table 4-4. Taxonomy and descriptive statistics (minimum (Min), maximum (Max), mean and standard deviation (*SD*)) of the OTUs with significant/suggestive associations with the five studied traits 146

List of abbreviations

ADG	average daily gain
ASV	amplicon sequence variant
BFT	back fat thickness
BLUP	best linear unbiased prediction
BV	breeding value
DFI	daily feed intake
ECR	energy conversion ratio
FCR	feed conversion ratio
GBLUP	genomic best linear unbiased prediction
GBV / GEBV	genomic estimated breeding value
GHG	greenhouse gases
HIF	heat increment of feeding
KR	Kleiber ratio
LD	linkage disequilibrium
ME	metabolisable energy
NE	net energy
OTU	operational taxonomic unit
QTL	quantitative trait loci
RDG	residual daily gain
REI	residual energy intake
RFI	residual feed intake
RIG	residual intake and gain
RMW	residual mid-test metabolic weight
SCFA	short chain fatty acids
SNP	single nucleotide polymorphism
ssGBLUP	single step genomic best linear unbiased prediction

1.

General introduction

1.1. Importance of feed in the pig industry

Changes in human population growth, income, lifestyle and eating habits have caused considerable rise in consumption of livestock products (FAO, 2013). Therefore, demand for livestock products has an increasing trend. Pig meat is one the most widely consumed sources of meat derived from domesticated animal species in the world and its production had a continuous increase from 1961 to 2018 (Figure 1-1).

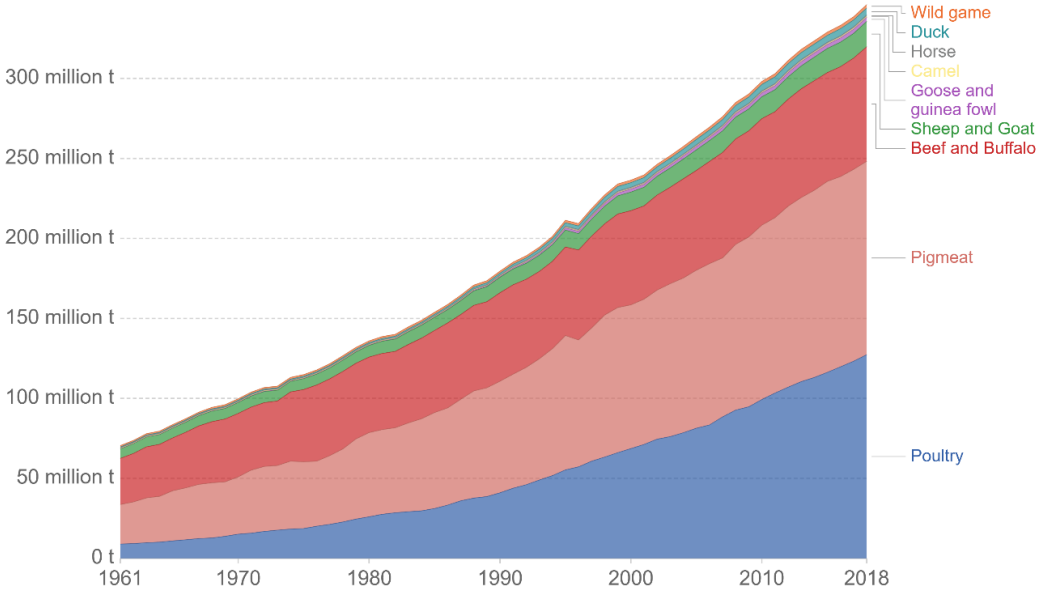


Figure 1-1. Meat production by type of livestock, World, 1961 to 2018

Source: UN Food and Agricultural Organization

Animal nutrition is one of the most important issues in animal husbandry. In other words, animal husbandry is based on proper nutrition of animals and the search for suitable feed. Proper nutrition is the feeding of livestock in a scientific manner so that besides the hygiene and housing costs, the maximum benefit can be obtained. Specifically, livestock nutrition accounts for more than half of the total production costs in industrialized countries. Therefore, by recognizing the feeding costs and having control on them with

considering the nutritional composition of the diet, more, faster and better livestock products can be achieved, and livestock can express their production potential.

1.1.1. Feeding costs in the pig industry

In an international comparison of conventional pig production costs conducted by the Wageningen Economic Research in 2018 (Hoste, 2017), total costs per kg carcass weight in some selected European countries ranged between €1.38 in Denmark to €1.88 in Italy, from which the feeding cost were €0.81 and €1.21, respectively. As the Figure 1-2 shows, the feeding costs per kg carcass weight in France and Netherland were equal to €0.85, which ranked second after Denmark. As this comparison confirmed, feeding costs in the European countries accounts for more than 50 % of the total production cost, and in France feeding costs reach to more than 59%. Therefore, decreasing the feeding costs is an outstanding challenge of the commercial pig production and would have a promising economic impact to improve the profit margins.

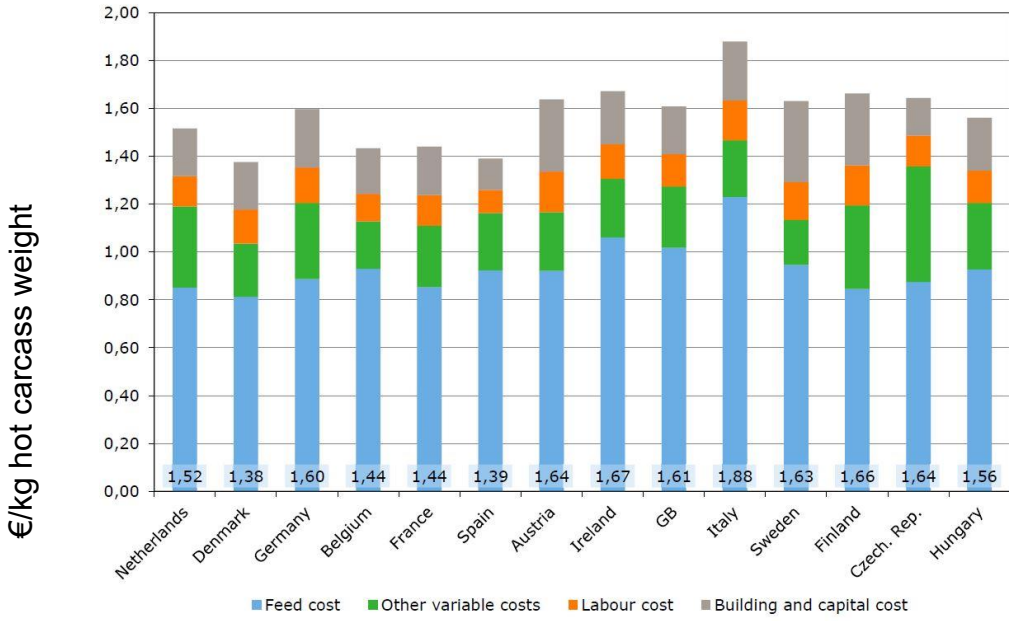


Figure 1-2. Cost of production (€/kg hot carcass weight), split into cost categories in selected EU countries on a closed cycle pig farm

Source: InterPIG/Wageningen Economic Research, year 2018.

1.1.2. Environmental impacts of nutrient excretion

Based on the estimates of the FAO, about one-third of total food supply of human is wasted or lost each year (FAO, 2013). This wastage mainly arises from opportunity missing and consequently imposes environmental impacts from food chains. It is also a relevant issue for the food consumption of farm animals, as reduced nutrient excretion would lead to less environmental impacts. According to the concept of sustainable production, environmental protection is an inevitable necessity for current and future generations, which is why today the environmental pollution crisis has become a global challenge and issue (Kupusovic et al., 2007). This challenge has had adverse environmental effects and consequences, such as pollution of water, air, soil, as well as endangering the health of humans and other living organisms. One of the most important of these problems is the increase in greenhouse gases (GHG) such as methane and carbon dioxide, which causes the continuous warming of the earth. Therefore, the production of environmentally friendly products and the processing of animal waste and scrap, including animal manure, can deal with the environmental problems caused by the release of these materials into the environment. Based on a global life cycle assessment conducted by FAO in 2013, the main GHG emissions source in the pig supply chains arise from feed production, that is contributing about 60 % of the total emission (MacLeod et al., 2013), whilst manure processing accounts for 27 % and the rest is related to post-farm processing and transportation of meat, direct and indirect energy use and enteric fermentation. The intensity of GHG emission has a strong relationship with the amount of natural resources used per unit of product (Fischedick et al., 2014). From the livestock breeding perspective, the efficiency of the use of feed by animals is a key controller of GHG emission (Herrero et al., 2013). Thus, improving the feed efficiency is an intervention to reduce emission of GHG at the animal and herd levels.

1.2. Definitions of feed efficiency and indicator traits

Feed efficiency is defined as the ratio of growth to feed consumption over a given period (Gaillard et al., 2020). Feed efficiency in terms of animal breeding quantifies how much an animal gains body weight with a given amount of feed, or its inverse, how much feed

it intakes for a given amount of body weight. Therefore, being more feed efficient means growing more or eating less compared to the other contemporary animals. In spite of its simplicity in calculation, feed efficiency has a complex nature, which involves variability in feed intake level, digestion and absorption of nutrients, metabolism and nutrient utilization, growth rate, body composition, physiological status of animals and many other environmental factors (Brito et al., 2020; Herd & Arthur, 2009; Li et al., 2016; Patience et al., 2015). The difficulties in measuring feed efficiency makes it necessary to use some indicator traits that are simultaneously accounting for feed intake and maintenance and growth requirements of animals. The first indicator trait used in livestock is feed conversion ratio (FCR), which is the traditional expression of feed efficiency and is defined as the ratio of average daily consumed feed (DFI: daily feed intake) to the average daily gain (ADG). Feed conversion ratio has been widely used to evaluate and improve feed efficiency for decades (Losinger, 1998; Pierozan et al., 2016). Nevertheless, improving feed efficiency through the direct selection for FCR is faced with difficulties because it is a ratio trait, which causes disproportional selection pressure on either DFI or ADG and difficulties in the prediction of response to selection (Gunsett, 1984). The other problem of FCR is that its distribution tends not to be normal, and can depend on the coefficient of variation of ADG, which also arises from the ratio nature of this trait (Aggrey & Rekaya, 2013; Atchley & Anderson, 1978; Yi et al., 2018). As an alternative indicator trait of feed efficiency, Koch et al. (1963) proposed residual feed intake (RFI) and applied it to beef cattle. Residual feed intake is defined as the difference between the observed average daily feed intake and that predicted from the average requirements for growth and maintenance of the animal, which is usually obtained using a multiple phenotypic regression model of DFI on metabolic body weight (for maintenance requirements), ADG and a backfat measurement (for production requirements), with fixed coefficients across animals. Because of its phenotypic independence from metabolic body weight and production traits, selection based on RFI would lead to better feed efficiency via decreased feed intakes while growth rate would be maintained or slightly reduced, whereas FCR leads to better efficiency via increased growth rates and slight decrease of feed intake, as shown in poultry by (Aggrey & Rekaya, 2013). Based on these outcomes, FCR is often qualified as “gross feed efficiency”, where RFI would indicate “net

feed efficiency" (Knap, 2009). In pigs, which differ more than poultry in the protein/lipid ratio of the body weight gain, leanness could also be differently affected by the choice of criteria (Saintilan et al., 2013). A common selection difficulty based on FCR is the ranking of two animals with same ratios (e.g., 2/1 and 4/2), whereas RFI, with taking into account of linear relationships between the components related to maintenance and production requirements, can deal with this issue (Aggrey & Rekaya, 2013).

Other feed efficiency metrics have been proposed, that are less known and used in practice to improve the feed efficiency. An instance is the residual daily gain (RDG) proposed by Koch et al. (1963). The RDG of a growing animal is defined as the residuals of the regression of ADG on FI. In contrast to RFI, the higher values of RDG are desirable, which indicate animals are gaining more weight than expected given their observed daily intake. The main disadvantage of RDG is its high dependence to ADG, which could confound associations with other performance traits (Ahola & Hill, 2012). Some of other metrics that are discussed by Calderón Díaz et al. (2017) are including ratio metrics such as energy conversion ratio (ECR), Kleiber ratio (KR) and relative growth rate (RGR) and residual metrics such as residual energy intake (REI), residual mid-test metabolic weight (RMW) and residual intake and gain (RIG).

1.3. Biological basis of feed efficiency

As mentioned above, numerous processes are involved in the variability of feed efficiency. They can be examined from the distribution of energy intake in different functions as represented in Figure 1-3. In pigs fed conventional diets, the main factors affecting the variation of feed efficiency have been identified after the digestion step (Noblet et al., 2013), despite the fact that digestive energy and nutrient losses represent 15 to 25% of the feed intake (Le Goff & Noblet, 2001) . The three main factors described in the literature are presented in more details in this section.

Heat dissipation: there are three components in heat dissipation: from basal metabolism (fasting heat production and maintenance processes), from feeding, and from activity. The digestive process produces additional natural heat known as the heat increment of feeding (HIF). The heat increment of feeding can be deduced from the

metabolisable energy (ME) to get the net energy (NE), which is the utilizable energy by the animal for maintenance and growth (Figure 1-3). Therefore, HIF is usually considered as an energetic loss and more feed efficient animals that are consuming less feed would have less energy expended as HIF (Herd & Arthur, 2009).

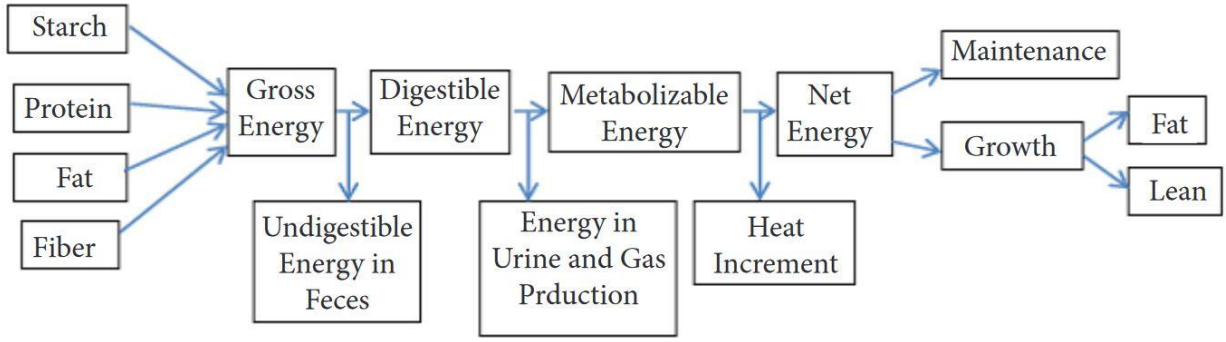


Figure 1-3. Dietary energy sources and energy use in the pig

Source: Euken (2012)

Activity: The physical activity of pigs is not part of the maintenance or growth requirements, and is another source of energy loss in the form of heat production. A study on growing pigs lines divergently selected for RFI showed that 14% of the feed intake difference between the lines is due to differences in activity level after 6 generations of selection (Meunier-Salaun et al., 2014). Some feeding behaviour traits, such as daily feeding time and daily number of visits to the feeder or feeding frequency, contribute to physical activity and have been shown to be significantly associated with the feed intake in pigs (Rauw et al., 2006) and in other species (poultry: Yan et al. (2019), sheep: Marie-Etancelin et al. (2019), cattle: Llonch et al. (2018)).

Composition of weight gain: More feed efficient animals are leaner and have less fat deposition than less feed efficient animals, which corresponds to negative correlations between FCR and leanness . A study on the effect of dietary energy on feed efficiency in pigs revealed that about 30 % to 35 % of the NE of diets is used for maintenance, 20 % to 25 % for protein gain, and the remaining 45 % to 50 % is used for lipid gain (Euken, 2012). This higher energy cost of fat deposition (~ 50 kJ of ME/g) than lean deposition (~ 40 kJ of ME/g) is due to the lower water content of fat tissue than lean

tissue (Rauw et al., 2017; van Milgen & Noblet, 2003). Nevertheless, higher lean tissue content is accompanied with higher maintenance requirements, because of the energy cost of protein turnover, and it has also been shown that fatter pigs produce less heat per unit of metabolic size than leaner pigs (Rauw et al., 2017; Sundstøl et al., 1979; Tess et al., 1984). Therefore, changing the body composition of animals toward more leanness would eventually lead to less demand for energy and reduced feed intake for the same amount of body mass. Nonetheless, feed efficient and lean animals can benefit more from the higher temperature of the environment than fat animals for maintenance and growth (Rauw et al., 2017), which can be considered in the management programs of the farms that feed efficiency is a goal trait. In addition, faster growth is also related with better feed efficiency, via a reduction of overall maintenance requirements to reach a given body weight, and increased protein deposition in earlier stages of growth.

Overall, under the concept of feed efficiency, animals that are able to direct a higher proportion of the net energy toward production are potentially more feed efficient (Brito et al., 2020). Therefore, any biological or environmental factor motivating this direction of energy would increase the feed efficiency of animals.

1.4. Means to improve the feed efficiency of pigs

1.4.1. Nutritional strategies

It is worth noting that improving feed efficiency is not simply formulating a diet with increasing energy concentration, as there is a low correlation between dietary energy concentration and feed efficiency if other nutrients are not accounted for (Patience et al., 2015). Thus, different nutritional strategies can be implemented to improve feed efficiency.

Energy and nutrient density: It has been proved that the energy level of the diet influences DFI and feeding time (Fracaroli et al., 2017; Patience, 2012). Increasing nutrient density of the diet relatively to the energy content is another nutritional strategy that can be adopted to improve feed efficiency. Since the energy and protein contents of the diet have high contribution to the carcass composition and quality, and account for

most of the diet cost, an optimized formulation of the diet in terms of energy and amino acids can help to improve feed efficiency (De Lange et al., 2001). In general, lysine is the main limiting amino acid in pig diets, and usually formulated diets should contain a specific level of lysine and crude protein (CP) to ensure an adequate supply of other amino acids. Such formulation may lead to an oversupply of CP in the diets, resulting in unnecessary excretion of nitrogen to the environment (Ball et al., 2013). Diets with reduced CP supplemented with crystalline amino acids are suggested to better deal with amino acids requirements of the animals, which can also control excessive protein intake and reduce nitrogen excretion (Ball et al., 2013; Le Bellego et al., 2001; Madrid et al., 2013; Tuitoek et al., 1997). Furthermore, reduction of nitrogen excretion means saving energy intake used to metabolize the excess protein, which can therefore be used for growth (Fracaroli et al., 2017). However, this energy might be used for fat deposition too, which needs more advanced formulation of the energy level of the diet (Fracaroli et al., 2017; Le Bellego et al., 2001; Madrid et al., 2013).

Diet form: several studies have shown that pelleted diets in comparison to mash or meal diets can significantly influence the improvement of the feed efficiency in pigs (Medel et al., 2004; Stark et al., 1993; Wondra et al., 1995). The reason of such improvement is the better digestibility of the pellet form as a result of processing steps, mainly temperature, heat and pressure, which provides more chemical and physical (particle size) availability of the nutrients (Noblet & van Milgen, 2004). Even the quality of the pellet is an important factor that affects the ratio of growth to feed intake (Stark et al., 1993). In addition, the type of feeder has a substantial role in the variation of ADG and DFI. Myers et al. (2013) in an investigation on the effects of feeder design on the growth performance of finishing pigs concluded that feeding pigs via feeders that allow the pigs to combine feed and water if they prefer (wet/dry Crystal Springs feeders) increase ADFI and ADG compared to conventional dry feeders. This increase was attributed to the fewer visits of the pigs with higher eating speed with wet/dry feeders (Bergstrom et al., 2012). Regarding the feeders type, feed spillage is a practical factor that can decrease feed efficiency. Gaillard et al. (2020) included the feed spillage in the equation of FCR as FCR = (feed intake + spillage) / pig growth. Conical semiautomatic feeders are suggested by

Pierozan et al. (2016) to reduce feed waste during feeding. Feeding pellets have also the advantage of less spillage (Vukmirović et al., 2017).

Diet digestibility: digestibility of the diet is associated with the fiber content, which usually is not digestible by endogenous digestive enzymes. Characterization of the fiber fraction in the livestock diets is usually based on content of crude fiber (CF), neutral detergent fiber (NDF) and acid detergent fiber (ADF). The traditional and still frequently applied measure is CF. However, it is less practical in the formulation of diets. NDF is indicator of plant's structural components such as cell walls, and more matured forages contain higher NDF. The ADF is an indicator of the least digestible components of plants and forages with low ADF values are desired. A fibrous diet usually enhances satiety and is related to lower NE values (Meunier-Salaun et al., 2001). However, since some indigestible fiber components are the main substrates for bacterial fermentation in the distal part of the gut, including some resources with dietary fibers in the diets is essential for the maintenance of the physiological functions in the gut (Wenk, 2001). The digestibility of the fibers differs between fiber sources and age of animals. In general, pigs can somewhat digest the dietary fibers, and this ability increases as they become more mature (Noblet & Le Goff, 2001). Noblet and Le Goff (2001) mentioned that the heat increment of dietary fiber could be used for thermoregulation or change the behaviour of pigs, as pigs fed with a fibrous diet tend to have less physical activity. In conclusion, a producer depending on the breeding goals and physiological status of the animals can consider all the properties of the fiber in the diet to improve the feed efficiency.

1.4.2. Genetic improvement of feed efficiency

1.4.2.1. Aspects of selection for feed efficiency

Genetic selection strategies to improve feed efficiency might be different when based on FCR or RFI, and depend on the breeding goals. Both traits require individual feed intake measurement, which can be costly. Even though FCR and RFI have high genetic and phenotypic correlations with each other (Table 1), in selection based on FCR, the economic aspect of feed would be more directly considered than in selection based on

RFI. In a selection program only based on FCR, the first focus would be to decrease the amount of feed intake per unit of body weight gain or vice versa, whereas a selection program based on RFI would only decrease feed intake while maintaining production and maintenance at the population average level. In theory, selection for RFI would be independent of production traits that are used to predict feed intake. For instance, it has been shown that after 9 generations of divergent selection for RFI, the two traits of ADG and BFT did not show significant changes (Gilbert et al., 2017). Therefore, the choice between selection based on FCR and RFI can be highly dependent on the source of variation of the feed intake of animals. A pig producer intending to select for the body composition of the animals while maintaining the growth rate level can adopt diverse strategy to improve the feed efficiency. Two main selection experiments on feed efficiency in Large White and Yorkshire growing pigs at INRAE and Iowa State University, respectively, have been shown successful development of two divergent lines that highlighted the biological responses of the selection for RFI (Cai et al., 2008; Gilbert et al., 2007). Some attributes of the LRFI animals resulting from several generations of divergent selection for RFI in the experiment conducted at INRAE were as in the following (Gilbert et al., 2007; Gilbert et al., 2017):

- Decreased technological meat quality,
- Increased nutritional requirements (g/MJ NE) and sensitivity to the density of diet nutrients / MJ NE
- Reduction in nitrogen and phosphorus excretion
- Reduced total amount of heat produced by unit of ME intake
- Reduced physical activity
- Non-significant changes in digestibility and robustness

Several studies on feed efficiency traits in different breeds (Table 1) have shown moderate heritability of FCR (ranged from 0.27 ± 0.05 to 0.45 ± 0.07) and low to moderate heritability for RFI (ranged from 0.12 ± 0.05 to 0.40 ± 0.04). The genetic correlation between the two traits are usually moderate to high, and ranges from 0.53 ± 0.07 to 0.88 ± 0.02.

Selection programs in commercial populations also showed improvement in feed efficiency of pig. For instance, a genetic progress evaluation of a selection program based on US terminal sire index showed −0.06 kg/kg genetic gain per year for FCR and 0.02 kg/d genetic gain per year for DFI in Duroc pigs with a generation interval of 1.5 year (Cheng et al., 2019). The genetic gains obtained from selection for FCR in commercial populations in France

Table 1-1. Heritability (h^2) and genetic correlations (r_G) of FCR and RFI in different pig breeds

Breed	h^2		r_G	
	FCR	RFI		
Duroc	0.39 ± 0.19	0.12 ± 0.05	0.75 ± 0.26	Sanchez et al. (2017)
Duroc	0.30 ± 0.04	0.38 ± 0.04	0.87 ± 0.04	Do et al. (2013)
Landrace	0.32 ± 0.05	0.36 ± 0.05	0.88 ± 0.02	Do et al. (2013)
Yorkshire	0.32 ± 0.05	0.40 ± 0.04	0.87 ± 0.03	Do et al. (2013)
French Landrace dam breed	0.35 ± 0.04	0.23 ± 0.03	0.53 ± 0.07	Saintilan et al. (2013)
Large White dam breed	0.30 ± 0.03	0.21 ± 0.03	0.52 ± 0.05	Saintilan et al. (2013)
Large White sire breed	0.30 ± 0.06	0.26 ± 0.06	0.69 ± 0.08	Saintilan et al. (2013)
Piétrain sire breed	0.40 ± 0.06	0.33 ± 0.06	0.85 ± 0.04	Saintilan et al. (2013)
Duroc	0.27 ± 0.05	0.38 ± 0.05	0.85 ± 0.13	Hoque et al. (2007)
Large White	0.45 ± 0.07	0.24 ± 0.03	0.71 ± 0.12	Gilbert et al. (2007)

1.4.2.2. Genomic selection in pig breeding programs

Recent theoretical and technological advances in using genomic information for the prediction of breeding values (BV) have provided means for more precise and feasible evaluation and selection of the animals for different traits. Prediction of BVs using genomic information (GBV) requires genotyping potential animals to be selected (candidates to selection), and genotyping training animals with performance traits from which the SNP effects are estimated (reference population). Based on the concept of the genetic progress (ΔG), the expected progress of a breeding program depends on the four factors of selection intensity (i), prediction accuracy (r), genetic variability (σ_g) and generation interval (L) as follows:

$$\Delta G = \frac{i.r.\sigma_g}{L}$$

Given this formula, the advantage of using genomic selection is related to the possible increase of prediction accuracy r and selection intensity i, and shortening of generation interval L, if it can allow selecting animals at younger stages than pedigree selection. However, unlike dairy cattle, the early use of young pigs as reproducers, the fact that most animals of a generation are candidates to selection and the short generation interval in pigs (maximum 2 years) limits the practical advantage of using genomic selection in pigs to the improvement of prediction accuracy (Tribout et al., 2011).

Besides the conventional factors affecting the BVs like heritability of traits, the accuracy of genomic predictions depends on the following specific factors (Clark et al., 2011; Daetwyler et al., 2012; Daetwyler et al., 2010; Druet et al., 2014; Habier et al., 2007; Meuwissen & Goddard, 2010):

- Number of animals in the training population
- Marker density
- Linkage disequilibrium (LD) between QTL and SNPs
- Effective population size
- Relatedness of selection candidates with individuals in the training dataset

- Genetic architecture of the traits: the distribution of QTLs effects and effective number of segments
- Imputation accuracy of marker genotypes
- Variance of relationships within the reference population.

Some of these factors interact with each other. For instance, the reason of using higher marker density panels is to better capture the extent of LD between markers and QTLs (Brito et al., 2011). The level LD can be affected by selection and effective population size, that result in higher levels of LD in livestock populations than in human (Khatkar et al., 2008). Even between livestock populations, it has been shown that the level of LD in pig populations is higher than in cattle populations (Veroneze et al., 2013).

The most outstanding advantage of genomic prediction is that with an adequate training set the prediction accuracy of BVs can be higher than traditional prediction methods (VanRaden et al., 2009). From an economic point of view, the gain in accuracy with the genomic prediction should be large enough to justify the expense of genotyping that is necessary for genomic evaluation (Abell et al., 2014). One of reasons of the lower field application of genomic selection in the pig industry, as compared to dairy cattle, is the low phenotyping cost of routine traits, even for later traits like reproduction traits, in comparison with the genotyping cost. In dairy cattle, pedigree selection was traditionally based to progeny testing, which generated very high selection accuracies at the expense of long generation intervals (~7 years) that corresponds to huge phenotyping costs. However, this statement in pigs is less relevant for the feed efficiency traits, as the cost of phenotyping is the main restricting factor of improvement programs for these traits, and often limits phenotyping to a sub-sample of the candidates to selection, thus reducing prediction accuracies.

The trait heritability is a determiner for the size of the training set as for traits with low heritability, larger training sets are required. In general, feed efficiency traits are moderately heritable (Table 1), which would help to optimize the number of animals in the training set. Therefore, genomic selection seems to be a promising strategy to achieve the desired prediction accuracy for feed efficiency traits.

Using imputation techniques to enhance the number of markers from a low-density panel to a high-density is a potential solution to deal with the high cost of genotyping (Dekkers et al., 2011; Habier et al., 2009). Depending on the species, genotyping with low-density panels can have lower cost than medium or high-density panels, and enables to increase the number of genotyped animals (Huang et al., 2012). An imputation with high accuracy can then provide accurate prediction of GBVs (Badke et al., 2014).

Pooling animals from different populations to construct the training set is another solution to reach high accuracy of genomic predictions. Building a pooled training set has substantial challenges, like LD differences and lack of strong relationships between sub-populations (Lund et al., 2014; Rezende et al., 2020). This strategy can however be considered for sub-populations across-countries, different breeds and different lines of limited size to achieve better prediction accuracies than with single populations alone (de Roos et al., 2009; Lund et al., 2014).

Depending on the prior assumption for the distribution of SNP effects, several statistical methods are available to obtain predictions of GBVs using SNP markers. Methods that are assuming a normal distribution and equal variances for all markers include snpBLUP or ridge regression BLUP (rrBLUP), genomic BLUP (GBLUP) and single step GBLUP (ssGBLUP) (Legarra et al., 2009; Meuwissen et al., 2001; VanRaden, 2008). The snpBLUP or rrBLUP method is based on the estimation of allelic effect of markers in the training set, followed by summing of these effects for the genotypes of the selection candidates. The GBLUP method is based on the use of mixed model equations with a genomic relationship matrix ($\mathbf{G}$). According to VanRaden (2008) the G matrix can be defined as following:

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum \mathrm{p_i}(1-\mathrm{p_i})}$$

The $\mathbf{Z}$ matrix is an n x m centralized matrix of genomic markers after deducting of $2(\mathrm{p_i} - 0.5)$, where $\mathrm{p_i}$ is the frequency of the major allele at locus i.

The ssGBLUP method incorporates the $\mathbf{G}$ matrix into the pedigree relationship matrix ($\mathbf{A}$) based on the decomposition of the $\mathbf{A}$ matrix into non-genotyped ($\mathbf{A_{11}}$) and

genotyped ($\mathbf{A_{22}}$) animals, so that it makes it possible to use all available phenotypic and pedigree information ($\mathbf{H}$ matrix) and to obtain GBVs for non-genotyped animals, which is not possible with the previous methods. The equation of the $\mathbf{H}$ matrix is written as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A_{11}} - \mathbf{A_{12}A_{22}^{-1}A_{21}} + \mathbf{A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21}} & \mathbf{A_{12}A_{22}^{-1}G} \\ \mathbf{GA_{22}^{-1}A_{21}} & \mathbf{G} \end{bmatrix}$$

The inverse of this matrix is easily obtained as:

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G^{-1}} - \mathbf{A_{22}^{-1}} \end{bmatrix}$$

The ssGBLUP is a promising way to improve the prediction accuracy of GBVs for feed efficiency traits that usually have limited phenotypic records. Finally, methods based on Bayesian approaches allow fitting different distributions of the SNP effects such as Bayes-A and Bayes-B, Bayes Cpi, Bayesian-Lasso and etc. (Calus, 2010; Gianola et al., 2009; Habier et al., 2011; Yi & Xu, 2008).

Altogether, in pig studies, prediction accuracies for growth and body composition traits are higher than feed efficiency traits (see Table 2 for examples of estimates from the literature), but the quantity of data available can differ between studies.

Table 1-2. Genomic prediction accuracy of growth and feed efficiency traits in some studies on pigs

Reference	Trait	Criterion	Breed	Accuracy
Guo et al. (2016)	ADG	$\mathrm{r(GEBV, y^*)/\sqrt{h^2}}$	Duroc	0.41
Guo et al. (2016)	BFT	$\mathrm{r(GEBV, y^*)/\sqrt{h^2}}$	Duroc	0.55
de Campos et al. (2015)	BFT	$\mathrm{r(GEBV, y^*)/\sqrt{h^2}}$	Duroc	0.61
Zhang et al. (2018)	DFI	$\mathrm{r(GEBV, y^*)/\sqrt{h^2}}$	Duroc	0.38 to 0.45
Christensen et al. (2012)	FCR	$\mathrm{r(GEBV, y^*)/\sqrt{h^2}}$	Duroc	0.16
Jiao et al. (2014)	RFI	$\mathrm{r(GEBV, y^*)/\sqrt{h^2}}$	Duroc	0.09

$\mathrm{y^*}$: adjusted phenotypes for fixed effects

In conclusion, increasing the number of animals in the training set has always been a challenge for genomic prediction. Therefore, any possibility to benefit from all available animals in the training set would provide higher prediction accuracy of GBVs and a more precise selection of animals.

1.4.3. Improving feed efficiency through digestion efficiency

The recent development of technologies in profiling the gastrointestinal tract (GIT) microbial communities have opened new opportunities for improving quantitative traits, specifically for feed related researches. This information is most often derived from partial sequencing of the bacterial 16S ribosomal RNA (rRNA) gene, a housekeeping gene in all bacteria (Woese, 1987). Sequencing the 16S rRNA gene has become a standard approach in bacterial taxonomic classification, due to its ease to generate phylogenetic information at high throughput (Wang et al., 2015). For this purpose, nine hypervariable regions (V1-V9) of the 16S rRNA gene can be targeted for sequencing. Sequences can then be clustered into 'Operational Taxonomic Units' (OTUs) based on their similarities, or each Amplicon Sequence Variant (ASV) can be analysed individually, which enables easy comparison between studies (Callahan et al., 2017). The OTUs (or ASV) are in fact the units that allow inferring the taxonomy of species present in the targeted biological samples. Identifying the taxonomy is facilitated by several reference databases, and can be used to propose hypotheses about the functionalities of the OTU. The counts of each OTU throughout the samples form a matrix called abundance table that is the basis of downstream analyses.

1.4.3.1. Influence of gut microbiome on feed efficiency

The interaction between a host animal and its GIT microbial community plays an essential dynamic role in the animal's vital processes including health status, physiological, immunological, nutritional and production processes (Mach et al., 2015). The microbiota is present throughout all parts of the GIT where bacteria are the predominant colonizing microorganisms in pigs (Stensland & Pluske, 2018). Numerous bacteria present in the

GIT of the pig are usually grouped into limited number of phyla, and mainly belong to the *Lactobacillus*, *Streptococcus*, *Peptococcus*, *Eubacterium*, *Clostridium*, *Bifidobacterium* and *Bacteroides* genera (Stensland & Pluske, 2018). The interaction between gut microbiota and host animal is mainly set via the degradation of fibrous resources, and production of substrate metabolites for the energy chain such as short, medium and long chain fatty acids, vitamins, biogenic amines and antimicrobials, through the fermentation of the nutrients by the bacteria (Broom & Kogut, 2018).

A healthy microbial composition, in addition to providing more resistance of the host to infectious diseases by stimulating the immune system and inhibiting pathogens, enables the host to effectively digest and absorb nutrients throughout the GIT (Backhed et al., 2005; Ducatelle et al., 2015; Stensland & Pluske, 2018). A healthy microbiota is generally characterised by high levels of diversity, as animals with a more diverse microbiota composition have more adaptability to the environmental changes and are more capable to deal with stressful periods in life, such as weaning (Stensland & Pluske, 2018). Commonness of the gut disorders in newly weaned pigs have directed the attention of researchers to this field in the last decades (Lalles et al., 2004). The link between microbiome and economic traits, and the improvement of its functions got increasing importance after the setting of new restrictive rules in European countries for using antibiotics and Zinc oxide (ZnO) as growth motivators in post-weaning diets (van Barneveld et al., 2018).

Several previous studies on pigs have investigated the link between the intestinal microbiota with growth, body composition and feed efficiency. Some important genera, such as *Bacteroides*, *Lactobacillus*, *Prevotella*, *Clostridium*, *Streptococcus*, *Roseburia*, *Coprococcus* and *Faecalibacterium*, have been reported to have association with ADG, BW, back fat, leanness and FE (Bergamaschi, Maltecca, et al., 2020; Han et al., 2017; Yang et al., 2016). However, there is not a full agreement about their effect on the performance traits, and conflicts on the reported associations do not allow getting a clear conclusion about important contributors. In fact, a large part of gut microbiota variation arises from differences in breeds, environmental conditions, diets, ages of pigs and the location of gut from which the samples are taken (Gardiner et al., 2020), and also heterogeneity of bioinformatics tools recruited for the analyses.

The association of different bacterial genera with feed efficiency in pigs is mainly described by their role in the degradation of carbohydrates and breaking down of plant-derived polysaccharides, which results in availability of short chain fatty acids (SCFA) as an energy source (Gardiner et al., 2020). Some of the genera that are reported to be more abundant in either ileum, faeces or caecum of the more feed-efficient pigs include *Christensenellaceae*, a polysaccharide degrader, *Treponema*, correlated with crude fibre digestibility, *Methanobrevibacter*, correlated with fibre digestibility, and *Actinobacillus,* a carbohydrate degrader and polysaccharide fermenter (McCormack et al., 2017; McCormack et al., 2019; Niu et al., 2015; Yang et al., 2017). The genera *Bacteroides* and *Clostridium*, by breaking down of N-glycan and degradation of polysaccharides, have been proposed as specifically associated with feed efficiency (McCormack et al., 2017; Yang et al., 2017). However, both genera have some pathogenic species that could cause the diversion of energy and nutrients towards the immune response rather than growth, resulting in negative correlation with feed efficiency (Songer & Uzal, 2005). Finally, he effect of some genera including *Ruminococcus*, *Butyricicoccus*, *Roseburia*, and *Lachnospiraceae* on feed efficiency is mores specifically described via the production of butyrate, which is one type of SCFA (McCormack et al., 2019; Quan et al., 2018; Quan et al., 2020; Tan et al., 2017; Vigors et al., 2020).

Another aspect of the effect of microbial bacteria on the feed efficiency could come from providing gut health and disease prevention by producing anti-inflammatory metabolites , which is proposed to explain some involvements of *Oscillibacter*, *Akkermansia* and *Lactobacillus* genera (McCormack et al., 2017; Quan et al., 2020; Valeriano et al., 2017; Vigors et al., 2020; Yang et al., 2017).

Finally, some negative effects of microbial bacteria on feed efficiency could be driven by the competing features of genera for nutrients with the host animal, as it has been mentioned for *Prevotella* and *Ruminococcus* genera that were mainly described as more abundant in less feed-efficient pigs (McCormack et al., 2019; Quan et al., 2020; Tan et al., 2017; Yang et al., 2017).

The reviewed literatures mostly focused on phenotypically contrasted groups of pigs for feed efficiency, that is an ideal to seek phenotypic relationships between feed

efficiency and GIT microbial community. The studies also subjected to heterogeneity in the microbiota objects studied, from the OTUs, which is the more complete data set, to families or genera that are restricted to the properly assigned OTUs. Nevertheless, biological interpretations at the family or genera levels are more sensible than at the OTU level. Overall, it seems that microbial components that are involved in the processing of nutrients, harvesting energy and those providing gut health and anti-inflammatory effects have positive associations with feed efficiency, and are enriched in the GIT of more feed-efficient animals, whereas pathogenic bacteria would be less abundant with negative effects. Based on this type of results, some authors have proposed phenotypic prediction of production traits using microbiota information (Mach et al, 2015), including feed efficiency (Le Sciellour et al., 2019; Verschuren et al., 2020), but with limited success.

1.4.3.2. Effect of host genetic on gut microbiome composition

Some studies revealed a substantial effect of the host genetic variance on the GIT microbiota composition of pigs, as in other species. Two main approaches are used, either by estimating heritability for different microbial taxa using classical animal mixed models (Camarinha-Silva et al., 2017), or by running genome-wide association studies, to identify genomic regions showing covariation with some microbiota components (Crespo-Piazuelo et al., 2019). Chen et al. (2018) reported 81 and 67 microbial taxa with heritability higher than 0.15 in fecal and cecum luminal samples, respectively, and identified candidate genes in genome wide association study (GWAS) that were mainly associated with metabolism, immunity functions and signal transduction. Similarly, Bergamaschi, Maltecca, et al. (2020) estimated non-zero heritabilities for OTUs of microbiome samples from faecal samples at weaning, at mid-test during the growth trial, and at the end of the growth trial. Therefore, part of the variation in GIT microbiome arises from genetic variation of the host, which could potentially be beneficial for future selection programs of feed efficiency if they also influence production traits. However, only few studies investigated the genetic relationships between feed efficiency and microbiome information in pigs (Bergamaschi, Tiezzi, et al., 2020; Camarinha-Silva et al., 2017). Besides GWAS and estimation of variance components for OTUs, estimating the microbiability m^2, which is the proportion of phenotypic variance explained by microbiota

information (Camarinha-Silva et al., 2017), can be beneficial to dissect this relationship and better understand the interplay between genetics and microbiota in the variability of production traits. As proposed by Camarinha-Silva et al. (2017), the microbiability can be obtained by computing a microbial relationship matrix $\mathbf{M}$ that structures the covariance between individuals due to the microbiota information. $\mathbf{M}$ is defined as $\mathbf{M} = \frac{\mathbf{Z_3 Z_3}'}{\mathrm{k}}$, where $\mathbf{Z_3}$ is a matrix with dimension of n x k, where n is the number of animals with microbiome information and m is the number of OTUs. Elements of the $\mathbf{Z_3}$ matrix are the standardized individual abundances of each OTU j for individual i, according to the following equation: $z_{3_{ij}} = \frac{\log(P_{ij}) - \overline{\log(P_j)}}{sd(\log(P_j))}$, where P_{ij} is the abundance of OTU j for individual i, and P_j is the vector of abundances of the jth OTU. Different formulas have been proposed for **M**, including the use of a 1-Jensen-Shannon distance between pairs of samples (Maltecca et al., 2019), Bray-Curtis distance matrix, or gene counts from metagenomics information rather than OTU abundances (in cattle: Difford et al, 2018, Ross et al, 2013). However, the Camarinha-Silva et al. (2017) computation is the most widely used in pig studies. With such approach, few estimates of microbiability for feed efficiency traits have been reported in the literature: 0.21 ± 0.14 from Camarinha-Silva et al. (2017) and 0.13 ± 0.10 from Weishaar et al. (2020) for FCR, and 0.45 ± 0.15 for RFI (Weishaar et al. (2020). Interestingly, some authors (Khanal et al., 2019) showed on backfat thickness that m² can be affected by age at sampling for some traits. Altogether, following the initial study by Camarinha-Silva et al. (2017) on 217 pigs, only very recent studies explored in a genetic framework the contribution of gut microbiota variability to feed efficiency traits in pigs.

1.5. Objectives

The main objective of the thesis was to investigate how genomic tools applied to the animal and its microbiota can contribute to improve selection for feed efficiency in pigs. The thesis relied on data collected in two pig lines during 10 generations of divergent selection for residual feed intake. Together with records on daily feed intake, growth, carcass composition and meat quality traits from at least two parities in all generations, tissue for pig DNA analyses were collected in every generation, and feces samples were collected from generations 9 and 10. Throughout the thesis, five production traits, including RFI, FCR, DFI, ADG and BFT, available on more than 1800 animals per line were investigated. Details on the population structure and development of the divergent lines are given in chapter 2.

To respond to the general objective of the thesis, the thesis was conducted in three chapters with the following specific objectives:

- In the second chapter, the main question was about the possible gains of accuracy for feed efficiency using genomic information: the focus will be on testing different genomic prediction scenarios that comprised animals from two different related lines. Such scenarios, because of the pedigree links between animals and consistent LD between sub-populations, could be more efficient than using across breeds or multi-breed genomic prediction to enhance prediction accuracy for this costly trait.

- In the third chapter, the genetic relationships between gut microbiota genera and feed efficiency and production traits were explored, with the objective to decipher whether the relationships between feed efficiency and gut microbiota had a genetic basis. Therefore, beside descriptive analyses of gut microbiota and comparisons between the two divergent lines, the heritability of genera and their genetic correlations with the five production traits will be presented and the possible biological bases discussed.

- In the fourth chapter, the objective was to decipher how microbiota globally contributes to variations of the production traits. To achieve this objective, the estimates of the microbiability were obtained for all traits, including or not the genetic relationship matrix in the models, and microbiome-wide association studies were run to find out which microbiome taxa have significant associations with the traits.

Following to the objectives of the MICROFEED project funded by the French National Research Agency that supported the thesis (ANR-16-CE20-0003), results of this study will be used to propose new genomic tools to jointly pilot the gut microbiota composition and the host genetic in terms of genetic selection for pig breeders, and in terms of nutrition and feeding for the feeding industry.

1.6. References

Abell, C. E., Dekkers, J. C., Rothschild, M. F., Mabry, J. W., & Stalder, K. J. (2014). Total cost estimation for implementing genome-enabled selection in a multi-level swine production system. *Genetics Selection Evolution*, *46*(1), 32. https://doi.org/10.1186/1297-9686-46-32

Aggrey, S. E., & Rekaya, R. (2013). Dissection of Koch's residual feed intake: implications for selection. *Poultry science*, *92*(10), 2600-2605. https://doi.org/10.3382/ps.2013-03302

Ahola, J. K., & Hill, R. A. (2012). Input factors affecting profitability: a changing paradigm and a challenging time. *Feed efficiency in the beef industry*, 7-19.

Atchley, W. R., & Anderson, D. (1978). Ratios and the statistical analysis of biological data. *Systematic Zoology*, *27*(1), 71-78.

Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *science*, *307*(5717), 1915-1920. https://doi.org/10.1126/science.1104816

Badke, Y. M., Bates, R. O., Ernst, C. W., Fix, J., & Steibel, J. P. (2014). Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. *G3 (Bethesda)*, *4*(4), 623-631. https://doi.org/10.1534/g3.114.010504

Ball, M. E., Magowan, E., McCracken, K. J., Beattie, V. E., Bradford, R., Gordon, F. J., Robinson, M. J., Smyth, S., & Henry, W. (2013). The effect of level of crude protein and available lysine on finishing pig performance, nitrogen balance and nutrient digestibility. *Asian-Australas J Anim Sci*, *26*(4), 564-572. https://doi.org/10.5713/ajas.2012.12177

Bergamaschi, M., Maltecca, C., Schillebeeckx, C., McNulty, N. P., Schwab, C., Shull, C., Fix, J., & Tiezzi, F. (2020). Heritability and genome-wide association of swine gut microbiome features with growth and fatness parameters. *Sci Rep*, *10*(1), 10134. https://doi.org/10.1038/s41598-020-66791-3

Bergamaschi, M., Tiezzi, F., Howard, J., Huang, Y. J., Gray, K. A., Schillebeeckx, C., McNulty, N. P., & Maltecca, C. (2020). Gut microbiome composition differences among breeds impact feed efficiency in swine. *Microbiome*, *8*(1), 110. https://doi.org/10.1186/s40168-020-00888-9

Bergstrom, J., Nelssen, J., Tokach, M., Dritz, S., Goodband, R., & DeRouchey, J. (2012). Effects of two feeder designs and adjustment strategies on the growth performance and carcass characteristics of growing–finishing pigs. *Journal of Animal Science*, *90*(12), 4555-4566.

Brito, F. V., Neto, J. B., Sargolzaei, M., Cobuci, J. A., & Schenkel, F. S. (2011). Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genetics*, *12*(1), 80. https://doi.org/10.1186/1471-2156-12-80

Brito, L. F., Oliveira, H. R., Houlahan, K., Fonseca, P. A. S., Lam, S., Butty, A. M., Seymour, D. J., Vargas, G., Chud, T. C. S., Silva, F. F., Baes, C. F., Canovas, A., Miglior, F., & Schenkel, F. S. (2020). Genetic mechanisms underlying feed utilization and implementation of genomic selection for improved feed efficiency in dairy cattle. *Canadian Journal of Animal Science*, *100*(4), 587-604. https://doi.org/10.1139/cjas-2019-0193

Broom, L. J., & Kogut, M. H. (2018). Gut immunity: its development and reasons and opportunities for modulation in monogastric production animals. *Anim Health Res Rev*, *19*(1), 46-52. https://doi.org/10.1017/S1466252318000026

Cai, W., Casey, D. S., & Dekkers, J. C. (2008). Selection response and genetic parameters for residual feed intake in Yorkshire swine. *Journal of Animal Science*, *86*(2), 287-298. https://doi.org/10.2527/jas.2007-0396

Calderón Díaz, J., Berry, D., Rebeiz, N., Metzler-Zebeli, B., Magowan, E., Gardiner, G., & Lawlor, P. (2017). Feed efficiency metrics in growing pigs. *Journal of Animal Science*, *95*(7), 3037-3046.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*, *11*(12), 2639-2643. https://doi.org/10.1038/ismej.2017.119

Calus, M. P. (2010). Genomic breeding value prediction: methods and procedures. *Animal*, *4*(2), 157-164. https://doi.org/10.1017/S1751731109991352

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S., & Bennewitz, J. (2017). Host Genome Influence on Gut Microbial Composition and Microbial

Prediction of Complex Traits in Pigs. *Genetics*, *206*(3), 1637-1644. https://doi.org/10.1534/genetics.117.200782

Chen, C., Huang, X., Fang, S., Yang, H., He, M., Zhao, Y., & Huang, L. (2018). Contribution of Host Genetics to the Variation of Microbial Composition of Cecum Lumen and Feces in Pigs. *Front Microbiol*, *9*, 2626. https://doi.org/10.3389/fmicb.2018.02626

Cheng, J., Newcom, D. W., Schutz, M. M., Muir, W. M., Cui, Q., Li, B., Zhang, H., & Schinckel, A. P. (2019). Selection for feed efficiency in Duroc pigs with the addition of daily feed intake data in a 2-stage selection procedure. *Applied Animal Science*, *35*(1), 20-29. https://doi.org/10.15232/aas.2018-01771

Christensen, O. F., Madsen, P., Nielsen, B., Ostersen, T., & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal*, *6*(10), 1565-1571. https://doi.org/10.1017/S1751731112000742

Clark, S. A., Hickey, J. M., & van der Werf, J. H. (2011). Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution*, *43*(1), 18. https://doi.org/10.1186/1297-9686-43-18

Crespo-Piazuelo, D., Migura-Garcia, L., Estelle, J., Criado-Mesas, L., Revilla, M., Castello, A., Munoz, M., Garcia-Casco, J. M., Fernandez, A. I., Ballester, M., & Folch, J. M. (2019). Association between the pig genome and its gut microbiota composition. *Sci Rep*, *9*(1), 8791. https://doi.org/10.1038/s41598-019-45066-6

Daetwyler, H. D., Kemper, K. E., van der Werf, J. H., & Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *Journal of Animal Science*, *90*(10), 3375-3384. https://doi.org/10.2527/jas.2011-4557

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, *185*(3), 1021-1031. https://doi.org/10.1534/genetics.110.116855

de Campos, C. F., Lopes, M. S., e Silva, F. F., Veroneze, R., Knol, E. F., Lopes, P. S., & Guimarães, S. E. (2015). Genomic selection for boar taint compounds and carcass traits in a commercial pig population. *Livestock Science*, *174*, 10-17.

De Lange, C., Marty, B., Birkett, S., Morel, P., & Szkotnicki, B. (2001). Application of pig growth models in commercial pork production. *Canadian Journal of Animal Science*, *81*(1), 1-8.

de Roos, A. P., Hayes, B. J., & Goddard, M. E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, *183*(4), 1545-1553. https://doi.org/10.1534/genetics.109.104935

Dekkers, J. C. M., Mathur, P. K., & Knol, E. F. (2011). Genetic Improvement of the Pig. *Genetics of the Pig, 2nd Edition*(Ed. 2), 390-425. https://doi.org/Doi 10.1079/9781845937560.0390

Do, D. N., Strathe, A. B., Jensen, J., Mark, T., & Kadarmideen, H. N. (2013). Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *Journal of Animal Science*, *91*(9), 4069-4079. https://doi.org/10.2527/jas.2012-6197

Druet, T., Macleod, I. M., & Hayes, B. J. (2014). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity (Edinb)*, *112*(1), 39-47. https://doi.org/10.1038/hdy.2013.13

Ducatelle, R., Eeckhaut, V., Haesebrouck, F., & Van Immerseel, F. (2015). A review on prebiotics and probiotics for the control of dysbiosis: present status and future perspectives. *Animal*, *9*(1), 43-48. https://doi.org/10.1017/S1751731114002584

Euken, R. M. (2012). Swine Feed Efficiency: Effect of Dietary Energy on Feed Efficiency.

Fischedick, M., Roy, J., Acquaye, A., Allwood, J., Ceron, J.-P., Geng, Y., Kheshgi, H., Lanza, A., Perczyk, D., & Price, L. (2014). Industry In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical Report.

Fracaroli, C., Perondi, D., dos Santos, L. S., da Silva, W. C., Veira, A. M., & Hauschild, L. (2017). Net energy levels of reduced crude protein, amino acid-supplemented diets for heavy pigs. *Livestock Science*, *205*, 43-49. https://doi.org/10.1016/j.livsci.2017.09.014

Gaillard, C., Brossard, L., & Dourmad, J. Y. (2020). Improvement of feed and nutrient efficiency in pig production through precision feeding. *Animal Feed Science and Technology*, *268*, 114611. https://doi.org/ARTN 114611

10.1016/j.anifeedsci.2020.114611

Gardiner, G. E., Metzler-Zebeli, B. U., & Lawlor, P. G. (2020). Impact of Intestinal Microbiota on Growth and Feed Efficiency in Pigs: A Review. *Microorganisms*, *8*(12), 1886. https://doi.org/10.3390/microorganisms8121886

Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., & Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics*, *183*(1), 347-363. https://doi.org/10.1534/genetics.109.103952

Gilbert, H., Bidanel, J. P., Gruand, J., Caritez, J. C., Billon, Y., Guillouet, P., Lagant, H., Noblet, J., & Sellier, P. (2007). Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. *Journal of Animal Science*, *85*(12), 3182-3188. https://doi.org/10.2527/jas.2006-590

Gilbert, H., Billon, Y., Brossard, L., Faure, J., Gatellier, P., Gondret, F., Labussiere, E., Lebret, B., Lefaucheur, L., Le Floch, N., Louveau, I., Merlot, E., Meunier-Salaun, M. C., Montagne, L., Mormede, P., Renaudeau, D., Riquet, J., Rogel-Gaillard, C., van Milgen, J., Vincent, A., & Noblet, J. (2017). Review: divergent selection for residual feed intake in the growing pig. *Animal*, *11*(9), 1427-1439. https://doi.org/10.1017/S175173111600286X

Gunsett, F. C. (1984). Linear Index Selection to Improve Traits Defined as Ratios. *Journal of Animal Science*, *59*(5), 1185-1193. <Go to ISI>://WOS:A1984TU28100009

Guo, X., Christensen, O. F., Ostersen, T., Wang, Y., Lund, M. S., & Su, G. (2016). Genomic prediction using models with dominance and imprinting effects for backfat thickness and average daily gain in Danish Duroc pigs. *Genetics Selection Evolution*, *48*(1), 67.

Habier, D., Fernando, R. L., & Dekkers, J. C. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, *177*(4), 2389-2397. https://doi.org/10.1534/genetics.107.081190

Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2009). Genomic Selection Using Low-Density Marker Panels. *Genetics*, *182*(1), 343-353. https://doi.org/10.1534/genetics.108.100289

Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, *12*(1), 186. https://doi.org/10.1186/1471-2105-12-186

Han, G. G., Lee, J. Y., Jin, G. D., Park, J., Choi, Y. H., Chae, B. J., Kim, E. B., & Choi, Y. J. (2017). Evaluating the association between body weight and the intestinal microbiota of weaned piglets via 16S rRNA sequencing. *Applied Microbiology and Biotechnology*, *101*(14), 5903-5911. https://doi.org/10.1007/s00253-017-8304-7

Herd, R. M., & Arthur, P. F. (2009). Physiological basis for residual feed intake. *Journal of Animal Science*, *87*(14 Suppl), E64-71. https://doi.org/10.2527/jas.2008-1345

Herrero, M., Havlik, P., Valin, H., Notenbaert, A., Rufino, M. C., Thornton, P. K., Blummel, M., Weiss, F., Grace, D., & Obersteiner, M. (2013). Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(52), 20888-20893. https://doi.org/10.1073/pnas.1308149110

Hoque, M. A., Suzuki, K., Kadowaki, H., Shibata, T., & Oikawa, T. (2007). Genetic parameters for feed efficiency traits and their relationships with growth and carcass traits in Duroc pigs. *Journal of Animal Breeding and Genetics*, *124*(3), 108-116. https://doi.org/DOI 10.1111/j.1439-0388.2007.00650.x

Hoste, R. (2017). *International comparison of pig production costs 2015: results of InterPIG*. Wageningen Economic Research.

Huang, Y., Hickey, J. M., Cleveland, M. A., & Maltecca, C. (2012). Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genetics Selection Evolution*, *44*(1), 25. https://doi.org/10.1186/1297-9686-44-25

Jiao, S., Maltecca, C., Gray, K., & Cassady, J. (2014). Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. *Journal of Animal Science*, *92*(6), 2377-2386.

Khanal, P., Maltecca, C., Schwab, C., Fix, J., & Tiezzi, F. (2019). Microbiability of meat quality and carcass composition traits in swine. *bioRxiv*, 833731.

Khatkar, M. S., Nicholas, F. W., Collins, A. R., Zenger, K. R., Cavanagh, J. A., Barris, W., Schnabel, R. D., Taylor, J. F., & Raadsma, H. W. (2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*, *9*(1), 187. https://doi.org/10.1186/1471-2164-9-187

Knap, P. (2009). Allocation of resources to maintenance. *Resource allocation theory applied to farm animal production*, *118*, 136.

Koch, R. M., Gregory, K. E., Chambers, D., & Swiger, L. A. (1963). Efficiency of Feed Use in Beef Cattle. *Journal of Animal Science*, *22*(2), 486-&. <Go to ISI>://WOS:A19633670B00025

Kupusovic, T., Midzic, S., Silajdzic, I., & Bjelavac, J. (2007). Cleaner production measures in small-scale slaughterhouse industry–case study in Bosnia and Herzegovina. *Journal of cleaner production*, *15*(4), 378-383.

Lalles, J. P., Boudry, G., Favier, C., Le Floc'h, N., Lurona, I., Montagne, L., Oswald, I. P., Pie, S., Piel, C., & Seve, B. (2004). Gut function and dysfunction in young pigs: physiology. *Animal Research*, *53*(4), 301-316. https://doi.org/10.1051/animres:2004018

Le Bellego, L., van Milgen, J., Dubois, S., & Noblet, J. (2001). Energy utilization of low-protein diets in growing pigs. *Journal of Animal Science*, *79*(5), 1259-1271. https://doi.org/10.2527/2001.7951259x

Le Goff, G., & Noblet, J. (2001). Comparative total tract digestibility of dietary energy and nutrients in growing pigs and adult sows. *Journal of Animal Science*, *79*(9), 2418-2427. https://doi.org/10.2527/2001.7992418x

Le Sciellour, M., Zemb, O., Hochu, I., Riquet, J., Gilbert, H., Giorgi, M., Billon, Y., Gourdine, J. L., & Renaudeau, D. (2019). Effect of chronic and acute heat challenges on fecal microbiota composition, production, and thermoregulation traits in growing pigs1,2. *Journal of Animal Science*, *97*(9), 3845-3858. https://doi.org/10.1093/jas/skz222

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, *92*(9), 4656-4663. https://doi.org/10.3168/jds.2009-2061

Li, F., Zhou, M., Ominski, K., & Guan, L. L. (2016). Does the rumen microbiome play a role in feed efficiency of beef cattle?1. *Journal of Animal Science*, *94*(suppl_6), 44-48. https://doi.org/10.2527/jas.2016-0524

Llonch, P., Somarriba, M., Duthie, C. A., Troy, S., Roehe, R., Rooke, J., Haskell, M. J., & Turner, S. P. (2018). Temperament and dominance relate to feeding behaviour and activity in beef cattle: implications for performance and methane emissions. *Animal*, *12*(12), 2639-2648. https://doi.org/10.1017/S1751731118000617

Losinger, W. C. (1998). Feed-conversion ratio of finisher pigs in the USA. *Prev Vet Med*, *36*(4), 287-305. https://doi.org/10.1016/s0167-5877(98)00094-4

Lund, M. S., Su, G., Janss, L., Guldbrandtsen, B., & Brondurn, R. F. (2014). Invited review: Genomic evaluation of cattle in a multi-breed context. *Livestock Science*, *166*, 101-110. https://doi.org/10.1016/j.livsci.2014.05.008

Mach, N., Berri, M., Estelle, J., Levenez, F., Lemonnier, G., Denis, C., Leplat, J. J., Chevaleyre, C., Billon, Y., Dore, J., Rogel-Gaillard, C., & Lepage, P. (2015). Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ Microbiol Rep*, *7*(3), 554-569. https://doi.org/10.1111/1758-2229.12285

MacLeod, M., Gerber, P., Mottet, A., Tempio, G., Falcucci, A., Opio, C., Vellinga, T., Henderson, B., & Steinfeld, H. (2013). *Greenhouse gas emissions from pig and chicken supply chains–A global life cycle assessment*. Food and Agriculture Organization of the United Nations.

Madrid, J., Martinez, S., Lopez, C., Orengo, J., Lopez, M. J., & Hernandez, F. (2013). Effects of low protein diets on growth performance, carcass traits and ammonia emission of barrows and gilts. *Animal Production Science*, *53*(2), 146-153. https://doi.org/10.1071/An12067

Marie-Etancelin, C., Francois, D., Weisbecker, J. L., Marcon, D., Moreno-Romieux, C., Bouvier, F., & Tortereau, F. (2019). Detailed genetic analysis of feeding behaviour in Romane lambs and links with residual feed intake. *Journal of Animal Breeding and Genetics*, *136*(3), 174-182. https://doi.org/10.1111/jbg.12392

McCormack, U. M., Curiao, T., Buzoianu, S. G., Prieto, M. L., Ryan, T., Varley, P., Crispie, F., Magowan, E., Metzler-Zebeli, B. U., Berry, D., O'Sullivan, O., Cotter, P. D., Gardiner, G. E., & Lawlor, P. G. (2017). Exploring a Possible Link between the Intestinal Microbiota and Feed Efficiency in Pigs. *Applied and Environment Microbiology*, *83*(15). https://doi.org/10.1128/AEM.00380-17

McCormack, U. M., Curiao, T., Metzler-Zebeli, B. U., Magowan, E., Berry, D. P., Reyer, H., Prieto, M. L., Buzoianu, S. G., Harrison, M., Rebeiz, N., Crispie, F., Cotter, P. D., O'Sullivan, O., Gardiner, G. E., & Lawlor, P. G. (2019). Porcine Feed Efficiency-Associated Intestinal Microbiota and Physiological Traits: Finding Consistent Cross-Locational Biomarkers for Residual Feed Intake. *mSystems*, *4*(4). https://doi.org/10.1128/mSystems.00324-18

Medel, P., Latorre, M. A., de Blas, C., Lazaro, R., & Mateos, G. G. (2004). Heat processing of cereals in mash or pellet diets for young pigs. *Animal Feed Science and Technology*, *113*(1-4), 127-140. https://doi.org/10.1016/j.anifeedsci.2003.08.005

Meunier-Salaun, M. C., Edwards, S. A., & Robert, S. (2001). Effect of dietary fibre on the behaviour and health of the restricted fed sow. *Animal Feed Science and Technology*, *90*(1-2), 53-69. https://doi.org/Doi 10.1016/S0377-8401(01)00196-1

Meunier-Salaun, M. C., Guerin, C., Billon, Y., Sellier, P., Noblet, J., & Gilbert, H. (2014). Divergent selection for residual feed intake in group-housed growing pigs: characteristics of physical and behavioural activity according to line and sex. *Animal*, *8*(11), 1898-1906. https://doi.org/10.1017/S1751731114001839

Meuwissen, T., & Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*, *185*(2), 623-631. https://doi.org/10.1534/genetics.110.116590

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829. <Go to ISI>://WOS:000168223400036

Myers, A. J., Goodband, R. D., Tokach, M. D., Dritz, S. S., DeRouchey, J. M., & Nelssen, J. L. (2013). The effects of diet form and feeder design on the growth performance

of finishing pigs. *Journal of Animal Science*, *91*(7), 3420-3428. https://doi.org/10.2527/jas.2012-5612

Niu, Q., Li, P., Hao, S., Zhang, Y., Kim, S. W., Li, H., Ma, X., Gao, S., He, L., Wu, W., Huang, X., Hua, J., Zhou, B., & Huang, R. (2015). Dynamic distribution of the gut microbiota and the relationship with apparent crude fiber digestibility and growth stages in pigs. *Sci Rep*, *5*, 9938. https://doi.org/10.1038/srep09938

Noblet, J., Gilbert, H., Jaguelin-Peyraud, Y., & Lebrun, T. (2013). Evidence of genetic variability for digestive efficiency in the growing pig fed a fibrous diet. *Animal*, *7*(8), 1259-1264. https://doi.org/10.1017/S1751731113000463

Noblet, J., & Le Goff, G. (2001). Effect of dietary fibre on the energy value of feeds for pigs. *Animal Feed Science and Technology*, *90*(1-2), 35-52. https://doi.org/Doi 10.1016/S0377-8401(01)00195-X

Noblet, J., & van Milgen, J. (2004). Energy value of pig feeds: effect of pig body weight and energy evaluation system. *Journal of Animal Science*, *82 E-Suppl*(suppl_13), E229-238. https://doi.org/10.2527/2004.8213_supplE229x

Patience, J. (2012). The influence of dietary energy on feed efficiency in grow-finish swine. In *Feed efficiency in swine* (pp. 101-129). Springer.

Patience, J. F., Rossoni-Serao, M. C., & Gutierrez, N. A. (2015). A review of feed efficiency in swine: biology and application. *J Anim Sci Biotechnol*, *6*(1), 33. https://doi.org/10.1186/s40104-015-0031-2

Pierozan, C. R., Agostini, P. S., Gasa, J., Novais, A. K., Dias, C. P., Santos, R. S. K., Pereira, M., Jr., Nagi, J. G., Alves, J. B., & Silva, C. A. (2016). Factors affecting the daily feed intake and feed conversion ratio of pigs in grow-finishing units: the case of a company. *Porcine Health Manag*, *2*(1), 7. https://doi.org/10.1186/s40813-016-0023-4

Quan, J., Cai, G., Ye, J., Yang, M., Ding, R., Wang, X., Zheng, E., Fu, D., Li, S., Zhou, S., Liu, D., Yang, J., & Wu, Z. (2018). A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios. *Sci Rep*, *8*(1), 4536. https://doi.org/10.1038/s41598-018-22692-0

Quan, J., Wu, Z., Ye, Y., Peng, L., Wu, J., Ruan, D., Qiu, Y., Ding, R., Wang, X., Zheng, E., Cai, G., Huang, W., & Yang, J. (2020). Metagenomic Characterization of Intestinal Regions in Pigs With Contrasting Feed Efficiency. *Front Microbiol*, *11*, 32. https://doi.org/10.3389/fmicb.2020.00032

Rauw, W. M., Mayorga, E. J., Lei, S. M., Dekkers, J. C. M., Patience, J. F., Gabler, N. K., Lonergan, S. M., & Baumgard, L. H. (2017). Effects of Diet and Genetics on Growth Performance of Pigs in Response to Repeated Exposure to Heat Stress. *Front Genet*, *8*, 155. https://doi.org/10.3389/fgene.2017.00155

Rauw, W. M., Soler, J., Tibau, J., Reixach, J., & Raya, L. G. (2006). The relationship between residual feed intake and feed intake behavior in group-housed Duroc barrows. *Journal of Animal Science*, *84*(4), 956-962. <Go to ISI>://WOS:000236658600021

Rezende, F. M., Haile-Mariam, M., Pryce, J. E., & Penagaricano, F. (2020). Across-country genomic prediction of bull fertility in Jersey dairy cattle. *Journal of Dairy Science*, *103*(12), 11618-11627. https://doi.org/10.3168/jds.2020-18910

Saintilan, R., Merour, I., Brossard, L., Tribout, T., Dourmad, J. Y., Sellier, P., Bidanel, J., van Milgen, J., & Gilbert, H. (2013). Genetics of residual feed intake in growing pigs: Relationships with production traits, and nitrogen and phosphorus excretion traits. *Journal of Animal Science*, *91*(6), 2542-2554. https://doi.org/10.2527/jas.2012-5687

Sanchez, J. P., Ragab, M., Quintanilla, R., Rothschild, M. F., & Piles, M. (2017). Genetic parameters and expected responses to selection for components of feed efficiency in a Duroc pig line. *Genetics Selection Evolution*, *49*(1), 86. https://doi.org/10.1186/s12711-017-0362-x

Songer, J. G., & Uzal, F. A. (2005). Clostridial enteric infections in pigs. *J Vet Diagn Invest*, *17*(6), 528-536. https://doi.org/10.1177/104063870501700602

Stark, C., Hines, R. H., Behnke, K. C., & Hancock, J. D. (1993). Pellet quality affects growth performance of nursery and finishing pigs.

Stensland, I., & Pluske, J. R. (2018). Use of probiotics and prebiotics in pig nutrition in the post-weaning period. In *Achieving sustainable production of pig meat Volume 2* (pp. 277-308). Burleigh Dodds Science Publishing.

Sundstøl, F., Standal, N., & Vangen, O. (1979). Energy Metabolism in Lines of Pigs Selected for Thickness of Backfat and Rate of Gain.

Tan, Z., Yang, T., Wang, Y., Xing, K., Zhang, F., Zhao, X., Ao, H., Chen, S., Liu, J., & Wang, C. (2017). Metagenomic Analysis of Cecal Microbiome Identified Microbiota and Functional Capacities Associated with Feed Efficiency in Landrace Finishing Pigs. *Front Microbiol*, *8*, 1546. https://doi.org/10.3389/fmicb.2017.01546

Tess, M. W., Dickerson, G. E., Nienaber, J. A., & Ferrell, C. L. (1984). The effects of body composition on fasting heat production in pigs. *Journal of Animal Science*, *58*(1), 99-110. https://doi.org/10.2527/jas1984.58199x

Tribout, T., Bidanel, J. P., Phocas, F., Schwob, S., Guillaume, F., & Larzul, C. (2011). La sélection génomique: principe et perspectives d'utilisation pour l'amélioration des populations porcines. 43. Journées de la Recherche Porcine,

Tuitoek, K., Young, L. G., de Lange, C. F., & Kerr, B. J. (1997). The effect of reducing excess dietary amino acids on growing-finishing pig performance: an elevation of the ideal protein concept. *Journal of Animal Science*, *75*(6), 1575-1583. https://doi.org/10.2527/1997.7561575x

Valeriano, V. D., Balolong, M. P., & Kang, D. K. (2017). Probiotic roles of Lactobacillus sp. in swine: insights from gut microbiota. *Journal of Applied Microbiology*, *122*(3), 554-567. https://doi.org/10.1111/jam.13364

van Barneveld, R., Hewitt, R., & D'Souza, D. (2018). Advances in understanding pig nutritional requirements and metabolism. In *Achieving sustainable production of pig meat Volume 2* (pp. 121-148). Burleigh Dodds Science Publishing.

van Milgen, J., & Noblet, J. (2003). Partitioning of energy intake to heat, protein, and fat in growing pigs. *Journal of Animal Science*, *81*(14_suppl_2), E86-E93.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414-4423. https://doi.org/10.3168/jds.2007-0980

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, *92*(1), 16-24. https://doi.org/10.3168/jds.2008-1514

Veroneze, R., Lopes, P. S., Guimaraes, S. E., Silva, F. F., Lopes, M. S., Harlizius, B., & Knol, E. F. (2013). Linkage disequilibrium and haplotype block structure in six commercial pig lines. *Journal of Animal Science*, *91*(8), 3493-3501. https://doi.org/10.2527/jas.2012-6052

Verschuren, L. M. G., Schokker, D., Bergsma, R., Jansman, A. J. M., Molist, F., & Calus, M. P. L. (2020). Prediction of nutrient digestibility in grower-finisher pigs based on faecal microbiota composition. *Journal of Animal Breeding and Genetics*, *137*(1), 23-35. https://doi.org/10.1111/jbg.12433

Vigors, S., JV, O. D., & Sweeney, T. (2020). Colonic microbiome profiles for improved feed efficiency can be identified despite major effects of farm of origin and contemporary group in pigs. *Animal*, *14*(12), 2472-2480. https://doi.org/10.1017/S1751731120001500

Vukmirović, Đ., Čolović, R., Rakita, S., Brlek, T., Đuragić, O., & Solà-Oriol, D. (2017). Importance of feed structure (particle size) and feed form (mash vs. pellets) in pig nutrition–A review. *Animal Feed Science and Technology*, *233*, 133-144.

Wang, X., Jordan, I. K., & Mayer, L. W. (2015). A phylogenetic perspective on molecular epidemiology. In *Molecular Medical Microbiology* (pp. 517-536). Academic Press.

Weishaar, R., Wellmann, R., Camarinha-Silva, A., Rodehutscord, M., & Bennewitz, J. (2020). Selecting the hologenome to breed for an improved feed efficiency in pigs—A novel selection index. *Journal of Animal Breeding and Genetics*, *137*(1), 14-22.

Wenk, C. (2001). The role of dietary fibre in the digestive physiology of the pig. *Animal Feed Science and Technology*, *90*(1-2), 21-33. https://doi.org/Doi 10.1016/S0377-8401(01)00194-8

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev*, *51*(2), 221-271. https://www.ncbi.nlm.nih.gov/pubmed/2439888

Wondra, K. J., Hancock, J. D., Behnke, K. C., Hines, R. H., & Stark, C. R. (1995). Effects of particle size and pelleting on growth performance, nutrient digestibility, and stomach morphology in finishing pigs. *Journal of Animal Science*, *73*(3), 757-763. https://doi.org/10.2527/1995.733757x

Yan, W., Sun, C., Wen, C., Ji, C., Zhang, D., & Yang, N. (2019). Relationships between feeding behaviors and performance traits in slow-growing yellow broilers. *Poultry science*, *98*(2), 548-555. https://doi.org/10.3382/ps/pey424

Yang, H., Huang, X., Fang, S., He, M., Zhao, Y., Wu, Z., Yang, M., Zhang, Z., Chen, C., & Huang, L. (2017). Unraveling the Fecal Microbiota and Metagenomic Functional Capacity Associated with Feed Efficiency in Pigs. *Front Microbiol*, *8*, 1555. https://doi.org/10.3389/fmicb.2017.01555

Yang, H., Huang, X., Fang, S., Xin, W., Huang, L., & Chen, C. (2016). Uncovering the composition of microbial community structure and metagenomics among three gut locations in pigs with distinct fatness. *Sci Rep*, *6*, 27427. https://doi.org/10.1038/srep27427

Yi, N., & Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics*, *179*(2), 1045-1055. https://doi.org/10.1534/genetics.107.085589

Yi, Z., Li, X., Luo, W., Xu, Z., Ji, C., Zhang, Y., Nie, Q., Zhang, D., & Zhang, X. (2018). Feed conversion ratio, residual feed intake and cholecystokinin type A receptor gene polymorphisms are associated with feed intake and average daily gain in a Chinese local chicken population. *J Anim Sci Biotechnol*, *9*(1), 50. https://doi.org/10.1186/s40104-018-0261-1

Zhang, C. Y., Kemp, R. A., Stothard, P., Wang, Z. Q., Boddicker, N., Krivushin, K., Dekkers, J., & Plastow, G. (2018). Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. *Genetics Selection Evolution*, *50*(1), 14. https://doi.org/ARTN 14

10.1186/s12711-018-0387-9

2.

Potential of the genomic information to improve selection for feed efficiency

2.1. Introduction

This chapter deals with the potential of the genomic information collected on the animals to improve selection for feed efficiency, with the hypothesis that information for this trait is usually scarcer than for other traits and FE could particularly benefit from genomic prediction technics. Most genomic predictions use a unique population that is split into a training and a validation set. However, how to enlarge the size or diversity of the training set, because of its potential high impact on prediction accuracy of GBVs, has always been a challenge for genomic prediction scenarios. Therefore, any possibility to benefit from more (diverse) animals in a training set could provide higher prediction accuracies of GBVs, and a more precise selection of animals. Besides, genomic prediction using genetically heterogeneous training sets could provide more flexibility when constructing the training sets for small populations. However, the literature shows quite heterogeneous results when combining populations for genomic prediction, and the aim of this chapter was to investigate the potential of genomic prediction for feed efficiency traits using training sets comprising animals from two related genetic lines. The GBVs were predicted using the single-step genomic best linear unbiased prediction method for six scenarios applied iteratively to the two genetically related lines (i.e. 12 scenarios) introduced before. The objective for all scenarios was to predict GEBV of pigs in the last three generations (~ 400 pigs, G7 to G9) of a given line. For each line, a control scenario was set up with a training set that included only animals from that line (target line).

For all traits, adding numerous animals from the other line, including early generations of selection, to the training set did not increase prediction accuracy compared to the control scenario. However, overall results showed that genomic prediction using a training set that included animals from genetically related lines can be as accurate as genomic prediction using a training set from the target population, depending on the relationship between the subsets. With combined reference sets, prediction accuracy increased for traits that were highly affected by selection, but biases also. These results provide insights into the design of reference populations, especially to initiate genomic selection in lines that are small, do not have a large number of historical samples and are developed simultaneously.

This chapter was published as a journal paper in Genetics, Selection, Evolution (DOI: 10.1186/s12711-020-00576-0). The supplementary material can be found in Appendix 2.1 (at the end of this chapter). In addition, early developments of the work were presented as a poster in the Gordon conference on Quantitative Genetics and Genomics in February 2019 in Luca, Italy (Appendix 2.2) and as an oral presentation at EAAP-2019 (Appendix 2.3).

2.2. Article I: The impact of training on data from genetically-related lines on the accuracy of genomic predictions for feed efficiency traits in pigs

Amir Aliakbari*, Emilie Delpuech, Yann Labrune, Juliette Riquet and Hélène Gilbert

GenPhySE, Université de Toulouse, INRAE, 31326 Castanet-Tolosan, France

Genetics Selection Evolution (2020), 52, 57

https://doi.org/10.1186/s12711-020-00576-0

2.2.1. Abstract

2.2.1.1. Background

Most genomic predictions use a unique population that is split into a training and a validation set. However, genomic prediction using genetically heterogeneous training sets could provide more flexibility when constructing the training sets in small populations. The aim of our study was to investigate the potential of genomic prediction of feed efficiency related traits using training sets that combine animals from two different, but genetically related lines. We compared realized prediction accuracy and prediction bias for different training set compositions for five production traits.

2.2.1.2. Results

Genomic breeding values (GEBV) were predicted using the single-step genomic best linear unbiased prediction method in six scenarios applied iteratively to two genetically related lines (i.e. 12 scenarios). The objective for all scenarios was to predict GEBV of pigs in the last three generations (~ 400 pigs, G7 to G9) of a given line. For each line, a control scenario was set up with a training set that included only animals from that line (target line). For all traits, adding more animals from the other line to the training set did not increase prediction accuracy compared to the control scenario. A small decrease in prediction accuracies was found for average daily gain, backfat thickness, and daily feed intake as the number of animals from the target line decreased in the training set. Including more animals from the other line did not decrease prediction accuracy for feed conversion ratio and residual feed intake, which were both highly affected by selection within lines. However, prediction biases were systematic for these cases and might be reduced with bivariate analyses.

2.2.1.3. Conclusions

Our results show that genomic prediction using a training set that includes animals from genetically related lines can be as accurate as genomic prediction using a training set from the target population. With combined reference sets, accuracy increased for traits

that were highly affected by selection. Our results provide insights into the design of reference populations, especially to initiate genomic selection in lines that are small, do not have a large number of historical samples and are developed simultaneously. This especially applies to poultry and pig breeding, as well as other crossbreeding schemes.

2.2.2. Background

Given the large economic impact of feed efficiency in the swine industry, its evaluation requires accurate estimation of breeding values (BV) and selection of animals (Patience et al., 2015). The most commonly used criterion to measure feed efficiency in livestock species is Feed Conversion Ratio (FCR) and is defined as feed intake per unit of live weight gain (Gaines et al., 2012). However, in 1963, residual feed intake (RFI) was introduced in cattle as an alternative criterion for feed efficiency (Koch et al., 1963). In general, FCR and RFI are highly genetically correlated (Hoque et al., 2007). Nevertheless, selection of animals based on FCR can be accompanied by undesirable correlated responses in other traits such as appetite (Ollivier et al., 1990; Pym & Nicholls, 1979), whereas selection for RFI is almost independent of these traits since RFI is feed intake adjusted for production trait by linear regression. Due to the high cost of measuring daily feed intake, and thus RFI and FCR [7], fewer phenotypic records are available, which reduces the accuracy of selection. Genomic selection has the potential to improve pig feed efficiency in some populations (Christensen et al., 2012; C. Y. Zhang et al., 2018). Recent advances in genomic evaluation methodologies, such as single-step genomic best linear unbiased prediction (ssGBLUP), enable more accurate evaluations in small populations. The ssGBLUP combines phenotypic, genotypic, and pedigree information in a single genomic evaluation of animals (Aguilar et al., 2010; Christensen & Lund, 2010; Legarra et al., 2009; Misztal et al., 2009). The number of animals in the reference population has been shown to affect the accuracy of genomic predictions (VanRaden et al., 2011). Multi-breed or admixed genomic evaluations have been proposed to increase the number of animals in reference sets for small populations (Carillier et al., 2014), resulting in increases in prediction accuracy in some cases (Lund et al., 2014). A study on multi-breed genomic evaluation using real data from Holstein and Jersey bulls showed that using a combined reference population resulted in comparable accuracies of

genomic estimated breeding values (GEBV) in purebred validation sets, or exceeded that achieved with a purebred reference population of the same breed (Ben J Hayes et al., 2009). Adding a smaller population, i.e. Brown Swiss, to a reference population of Holstein and Jersey bulls resulted in slight increases in accuracy of predictions when breeds were considered as a single, joint population, while slight increases in accuracy were also observed if the breeds were treated as genetically related traits (Olson et al., 2012). Simulation studies with mixed reference populations also showed increases in prediction accuracy. A simulation study on genomic prediction across multiple populations in cattle showed that adding relatively few individuals from another population to a training set substantially increased the accuracy of predictions in the first population, regardless of the heritability (h^2) or marker density (de Roos et al., 2009). Another simulation study reported that genomic predictions using a combined versus a single reference population increased the accuracy of genomic predictions by 25%, with traits with a lower heritability benefiting more from the combination of populations (S.-Y. Zhang et al., 2018). However, using a combined reference population can be challenging if relationships between populations are absent: allele frequencies at the marker and/or causal loci, or causal variants themselves, can differ between populations, (Carillier et al., 2014; Lund et al., 2014). Another limitation for across-breed genomic prediction is the inconsistency of linkage disequilibrium (LD) between markers and quantitative trait loci (QTL) between breeds, which is one of the assumptions of most genomic prediction models (Ben J Hayes et al., 2009).

Given the presence of (ancestral) relationships between animals and the greater consistency of LD between genetically related lines within a breed than between breeds that have been separated for decades, using a multi-line reference population may be more beneficial than using a multi-breed reference population (Lund et al., 2014). However, the changes in allele frequency since separation of the lines may still represent a challenge for using a multi-line reference population (Fangmann et al., 2015). To the best of our knowledge, the use of a multi-line genomic evaluation strategy in small, related lines using real data has not been studied, despite the existence of numerous related lines worldwide. Our hypothesis was that, in small porcine populations with few available ancestral samples, i.e. cannot build large reference populations, including information

from a genetically related line in the training population could provide similar prediction accuracy as a within-line training population. Therefore, we explored reference populations with different structures that combined data from two lines that descended from a common origin, and compared the prediction accuracy obtained with that obtained when only information from the target line was used for training.

2.2.3. Methods

2.2.3.1. Population and data structure

The data were collected during a selection experiment that was conducted at INRAE (UE GenESI, Surgères, France, https://doi.org/10.15454/1.5572415481185847E12) on French Large White pigs. Two lines were established by nine generations of divergent selection for RFI from 2000 to 2015 (Gilbert et al., 2017). The G0 generation resulted from the mating of 30 boars and 30 gilts from generation F0 using artificial insemination. Among the G0 animals, 116 boar candidates for selection from all 30 litters were tested for RFI to select six extreme founder boars for each line (LRFI: low RFI, and HRFI: high RFI). The two lines were initiated by mating the selected boars to about 35 random G0 gilts per line. Inbreeding was minimized at each generation. The development of each line continued with the selection of six boars out of 96 tested candidates in each generation from G1 to G9. In each generation, at least one additional parity was produced to evaluate correlated responses to selection for production traits on both females and castrated males (henceforth referred to as response animals). Selection candidates were evaluated for RFI from 35 to 95 kg of body weight (BW), and response animals were evaluated from 10 weeks of age until slaughter (105 kg BW until G5 and 115 kg BW from G6 onwards). Animals were raised in four pens per batch and at least four batches per generation. Test pens were equipped with single-place electronic feeders ACEMA64 (ACEMO, France). Animals were offered ad libitum access to a pelleted diet based on cereals and soya bean meal containing 10 MJ net energy (NE)/kg and 160 g CP/kg, with a minimum of 0.80 g digestible Lys/MJ NE. In each generation, boars were selected based on a fixed RFI selection index that was established from pre-computed phenotypic correlations between daily feed intake (DFI, g/d) and average daily gain (ADG, g/d)

between 35 and 95 kg BW, and live backfat thickness (BFT, mm) at 95 kg BW (Gilbert et al., 2007), as RFI = DFI − 1.06 × ADG − 37 × BFT. The average metabolic BW (AMBW) was the same for all selection candidates and therefore excluded from the selection index equation. Selection candidates had records for feed intake, body weight, and live body composition traits. In addition to these phenotypes, gilts and castrated males had records for carcass composition traits (Gilbert et al., 2007). For the present study, RFI, FCR, DFI, ADG and BFT were analyzed. These traits were available for both selection candidates and response animals. The number of observations for the five traits for each line are in Table 1. RFI of selection candidates was computed between 35 and 95 kg BW as the residual of a multiple linear regression of DFI on the traits included in the selection index. For gilts and castrated males from the correlated response batches, RFI was estimated from 10 weeks of age to slaughter as the residual of a multiple linear regression of DFI on AMBW, ADG from 10 weeks of age to slaughter, carcass BFT (carcBFT), and lean meat content (LMC; computed from cut weights) at slaughter. AMBW was included to account for maintenance requirements and the other traits were included to account for production requirements. (Gilbert et al., 2017). Fixed effects included in the regression model to compute RFI of response animals were sex, pen size, contemporary group and BW at the beginning of the test. Complete pedigree information was collected from F0 to G9, plus up to 10 generations of ancestors, and contained 7046 animals (Table 1).

2. Potential of the genomic information to improve selection for feed efficiency

Table 2-1. Numbers of animals in the pedigree and data structure

	Ancestors	F0	G0	HRFI										
				G0	G1	G2	G3	G4	G5	G6	G7	G8	G9	Total
Pedigree	159	67	104	48	216	297	277	260	270	795	474	292	280	3209
Pedigree only				1	2	89	78	62	68	352	149	5	0	806
Pedigree and genotype only				41	41	42	44	36	47	40	35	42	91	459
ADG														
Phenotype only				0	167	160	149	156	149	304	194	148	93	1520
Phenotype and genotype				6	6	6	6	6	6	71	73	66	92	338
Missing				0	0	0	0	0	0	28	23	31	4	86
BFT														
Phenotype only				0	167	160	149	156	149	237	176	62	84	1340
Phenotype and genotype				6	6	6	6	6	6	71	73	66	92	338
Missing				0	0	0	0	0	0	95	41	117	13	266
DFI														
Phenotype only				0	166	160	149	156	149	263	182	138	93	1456
Phenotype and genotype				6	6	6	6	6	6	71	73	66	92	338
Missing				0	1	0	0	0	0	69	35	41	4	150
FCR														
Phenotype only				0	166	160	148	156	149	263	182	138	93	1455
Phenotype and genotype				4	6	6	6	6	6	71	73	66	92	336
Missing				2	1	0	1	0	0	69	35	41	4	153
RFI														
Phenotype only				0	164	159	146	156	143	185	147	56	80	1236
Phenotype and genotype				6	6	6	6	6	6	71	73	66	92	338
Missing				0	3	1	3	0	6	147	70	123	17	370
	Ancestors	**F0**	**G0**	**LRFI**										
				G0	**G1**	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**	**G9**	**Total**
Pedigree	159	67	104	46	203	303	314	327	357	826	481	344	280	3481
Pedigree only				0	1	98	100	107	130	337	132	8	0	913
Pedigree and genotype only				40	35	40	41	43	43	48	55	48	93	486
ADG														
Phenotype only				0	161	159	167	171	178	359	211	203	95	1704
Phenotype and genotype				6	6	6	6	6	6	74	73	74	90	347
Missing				0	0	0	0	0	0	8	10	11	2	31
BFT														
Phenotype only				0	161	159	167	171	178	284	206	105	86	1517
Phenotype and genotype				6	6	6	6	6	6	74	73	74	90	347
Missing				0	0	0	0	0	83	15	109	1	1	218
DFI														
Phenotype only				0	160	159	167	171	178	316	206	194	95	1646
Phenotype and genotype				6	6	6	6	6	6	74	73	74	90	347
Missing				0	1	0	0	0	0	51	15	20	2	89
FCR														
Phenotype only				0	159	159	167	171	178	316	208	195	95	1648
Phenotype and genotype				6	6	6	6	6	74	73	74	90		347
Missing				0	2	0	0	0	0	51	13	19	2	87
RFI														
Phenotype only				0	160	158	161	171	173	230	165	101	80	1399
Phenotype and genotype				6	6	6	6	6	6	74	73	74	90	347
Missing				0	1	1	6	0	5	137	56	113	17	336

HRFI high RFI line, *LRFI* low RFI line, *Ancestors* animals before the base generation, *F0* base generation, *G0 to G9* generations of selection 0 to 9, *RFI* residual feed intake, *ADG* average daily gain, *FCR* feed conversion ratio, *DFI* daily feed intake, *BFT* backfat thickness

2.2.3.2. Combining and standardizing traits

Preliminary analyses on the five traits showed high genetic correlations between similar traits measured in selection candidate and response animals (> 0.80 ± 0.11, except 0.75 ± 0.08 between live BFT and carcass BFT). Therefore, to increase the amount of information, corresponding traits in selection candidate and response animals were combined for further analyses. Animals differed in age and BW when measurements were taken. Therefore, for each trait, records from selection candidates were standardized to the variance of the corresponding trait in the response animals as:

$$y_{Rij} = \frac{y_{sij}}{\sigma_{si}}\sigma_{Ri},$$

where y_{Rij} is the standardized trait i (i = 1 … 5) for selection candidate j, y_{sij} is the record of trait i measured on animal j, σ_{si} is the phenotypic standard deviation of trait i measured on selection candidates, and σ_{Ri} is the phenotypic standard deviation of trait i measured on females and castrated males in the response batches. Descriptive statistics of these traits are in Table 2.

Table 2-2. Descriptive statistics of the data for the studied traits in the HRFI and LRFI lines

Line	Trait	Number of records	Minimum	Maximum	Average	Coefficient of variation
HRFI	ADG	1868	0.44	1.07	0.76	11.03
	BFT	1687	9.67	49.27	27.33	26.62
	DFI	1802	1.37	3.20	2.18	12.54
	FCR	1799	2.13	3.81	2.8	9.26
	RFI	1581	- 0.29	0.86	0.05	–
LRFI	ADG	2053	0.45	1.06	0.76	10.69
	BFT	1866	10.00	44.63	26.45	24.60
	DFI	1995	1.05	2.92	2.01	12.91
	FCR	1997	1.72	3.70	2.60	9.11
	RFI	1748	- 0.56	0.46	- 0.04	–

HRFI high RFI line, *LRFI* low RFI line, *ADG* average daily gain (kg/day), *BFT* backfat thickness (mm), *DFI* daily feed intake (kg/day), *FCR* feed conversion ratio (kg/kg), *RFI* residual feed intake (kg/day)

2.2.3.3. Single nucleotide polymorphism (SNP) genotyping data and imputation

SNP genotyping data were available for all selected boars and their mates from G0 to G9, additional pigs from response batches of G6 to G8, and all selection candidates in G9. In total, 1647 animals had SNP genotypes, of which 286 animals were genotyped with the Porcine SNP60v2 BeadChip (Illumina) (64,232 SNPs) and 1361 animals with the GGP Porcine HD Array (Illumina) (68,516 SNPs). Genotype quality control excluded SNPs with a call rate lower than 95%, individuals with a call rate lower than 90%, SNPs that were not in Hardy-Weinberg equilibrium ($p < 10^{-10}$), SNPs with minor allele frequency lower than 0.01, and individuals with parent-offspring incompatibility (e.g., opposite homozygotes) with at least one parent. The PLINK software was used for SNP and individual genotype quality control (Purcell et al., 2007). SNPs on the sex chromosomes were removed. After quality control of each SNP chip dataset, the SNPs present in each panel were imputed to the alternative panel using the FImpute software (Sargolzaei et al., 2014) in a single step. The two SNP chips shared 42,800 SNPs. The number of genotyped animals retained after imputation was 1643, and the final genotype dataset contained 64,233 informative SNPs. Thus, all animals had equal genotypic information. Genotypes were coded as 0, 1, or 2 for later calculation of the genomic relationship matrix. The number of animals with genotype data per generation and line is in Table 2-1.

2.2.3.4. Model and analyses

Predictions obtained with BLUP are based on the assumption of no genetic differences between subpopulations (Careau et al., 2013; Hadfield et al., 2010). Therefore, to account for selection in our dataset, all genetic and genomic analyses were carried out with bivariate approaches. All other five traits were individually paired with the selection index in two-trait model analyses. By including the selection criterion, the analyses of other traits are conditioned based on all the information that was used for selection (Fernando & Gianola, 1990; Henderson, 1990; Sorensen et al., 2001).

Preliminary analyses were carried out using a general linear model in R (glm procedure) to evaluate the significance ($p < 0.05$) of fixed environmental sources of variation. The significant fixed factors included pen size (5 levels: 8, 9, 10, 11, 12 pigs per

pen), herd of birth (2 levels), sex (3 levels), and contemporary groups (CG, 99 levels). BW at slaughter was fitted in the model as a covariate only for BFT. CG were defined as animals born in the same week and raised in the same enclosure. Litter was fitted as a random environmental source of variation and its significance at the 5% level was determined using a likelihood ratio test.

The genetic analyses were performed using the AIREMLF90 and BLUPF90 software (Misztal et al., 2018) for the BLUP and ssGBLUP methods, respectively. Prior to ssGBLUP evaluations, the variance components of the traits were obtained using the restricted maximum likelihood algorithm implemented in AIREMLF90. These analyses were performed using all available data and only the full pedigree relationship matrix ($\mathbf{A}$). Variance components were estimated with the bivariate animal mixed model as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_1 a} + \mathbf{Z_2 l} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of observations for the index and one of the five studied traits, $\mathbf{b}$ is the vector of fixed effects (described above), $\mathbf{a}$ is the vector of additive genetic effects, $\mathbf{l}$ is the vector of litter effects, and $\mathbf{e}$ is the vector of random residuals. $\mathbf{X}$, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices for $\mathbf{b}$, $\mathbf{a}$, and $\mathbf{l}$, respectively. Distributions assumed for the random terms are $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G_0} \otimes \mathbf{A})$, $\mathbf{l} \sim N(\mathbf{0}, \mathbf{R_l} \otimes \mathbf{I})$, and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R_e} \otimes \mathbf{I})$, where $\mathbf{G_0}$ is a 2×2 symmetric (co)variance matrix of direct additive genetic effects, and $\mathbf{R_l}$ and $\mathbf{R_e}$ are 2×2 symmetric (co)variances matrices of litter and residual effects, respectively. $\mathbf{I}$ denotes the identity matrix.

Genomic breeding values were estimated using ssGBLUP with the same models in the BLUPF90 software, with the previously estimated (co)variances and using the $\mathbf{H}$ matrix, which is a combined relationship matrix of the $\mathbf{A}$ matrix and marker-based relationship matrix ($\mathbf{G}$) of genotyped animals (Aguilar et al., 2010; Legarra et al., 2009). The $\mathbf{G}$ matrix was constructed and scaled by $2\sum\{p_i(1-p_i)\}$, where p_i is the frequency of the second allele at locus i, following VanRaden (2008). Computation of the $\mathbf{H}$ matrices used outputs of BLUPF90 ($\mathbf{G}$) and the full $\mathbf{A}$ matrix, which was obtained using the AGHmatrix R package (Amadeu et al., 2016). In all scenarios, $\mathbf{G}$ had similar average diagonal elements as the pedigree relationship matrix for the genotyped animals ($\mathbf{A_{22}}$).

2.2.3.5. Scenarios

Two symmetric series of six scenarios, one for each line, were defined for genomic prediction. An overview of the scenarios is shown in Figure 2-1. In all scenarios, genotyped animals of the last three generations (G7 to G9, 433 pigs for the LRFI and 399 pigs for the HRFI line) were considered for validation in a given line (target line), and their information was removed from the training dataset.

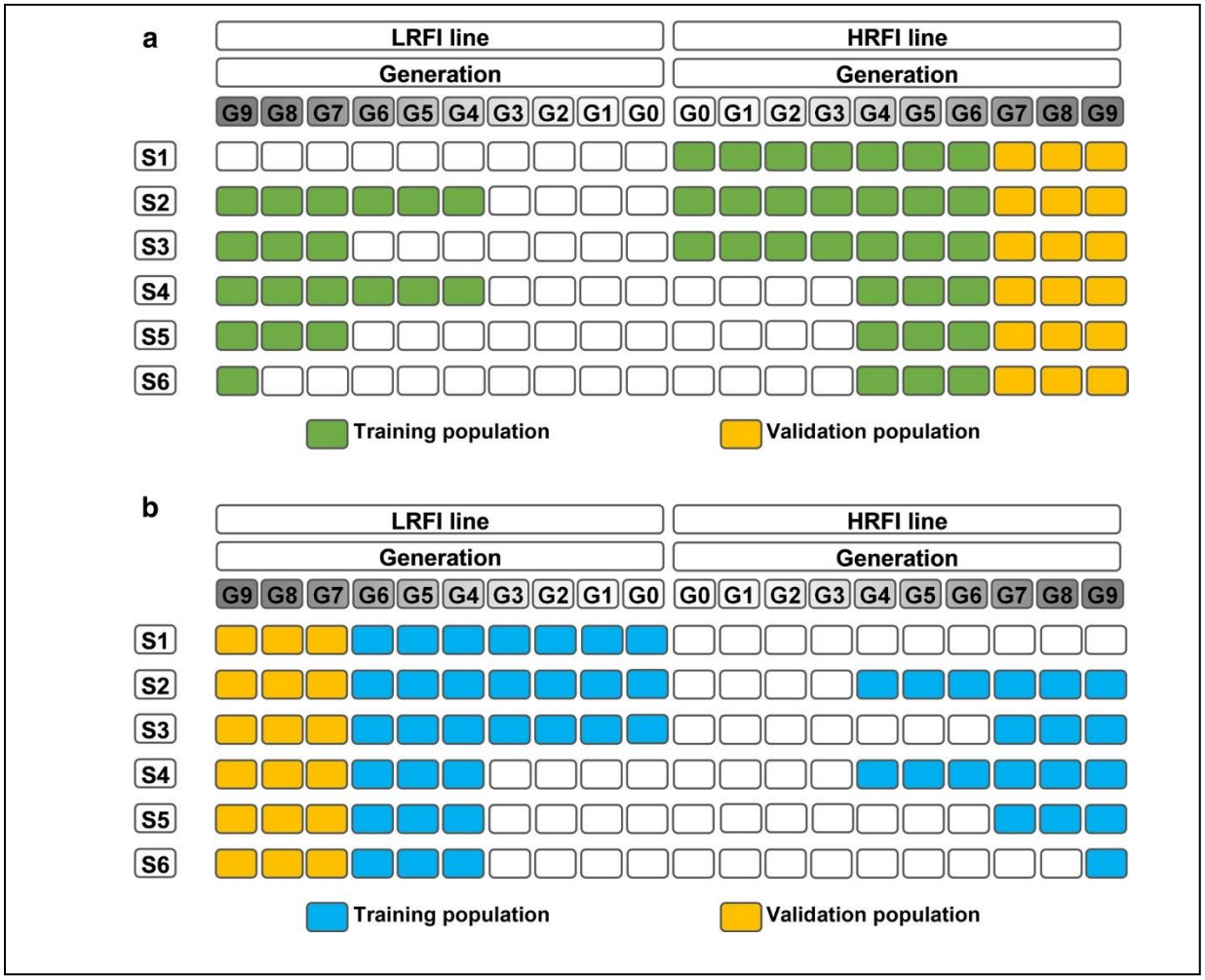


Figure 2-1. Design of scenarios to predict validation animals in HRFI (a) and LRFI (b) lines

The training sets were structured based on which generations and line were used. Scenario 1 comprised only animals from the target line and was the control scenario since it represented a routine genomic prediction design where all data would be available from the same line. All other scenarios were compared to this control scenario to evaluate which combination of training populations from the two lines achieved a prediction accuracy similar to the control scenario. Scenarios 2 and 3 included the training set of scenario 1 and additionally, either the animals from G4 to G9 (scenario 2), or G7 to G9 (scenario 3) of the other line.

For scenarios 4 to 6, animals from the target line in the training set were limited to the three generations nearest to the validation set (G4 to G6). In scenarios 4 and 5, the contribution to the training set of the animals from the other line was as in scenario 2 (G4 to G9) and scenario 3 (G7 to G9), respectively. For scenario 6, the number of animals in the training set was equal to that of scenario 1 and only animals from the G9 generation of the other line. Performance data of animals from the generation and line combinations that did not contribute to the training or validation sets were removed from the analysis, but their pedigree information was kept in order to trace relationships back to the founding generation. For example, phenotypes and genotypes of animals from G0 to G3 of both lines were removed for scenario 4, since they were not part of the training or validation sets. The number of genotyped animals in the training and validation sets for the 12 scenarios are in Table 2-3.

Table 2-3. Number of genotyped animals in the training and validation sets for the six scenarios for the HRFI and LRFI validation sets

	HRFI		LRFI	
	Training	**Validation**	**Training**	**Validation**
Scenario 1	398	399	400	433
Scenario 2	1051	399	1005	433
Scenario 3	831	399	799	433
Scenario 4	859	399	825	433
Scenario 5	639	399	619	433
Scenario 6	389	399	403	433

HRFI high RFI line, *LRFI* low RFI line

2.2.3.6. Accuracy and bias of genomic predictions

Usually the correlation between the vector of estimated breeding values ($\mathbf{EBV}$) to be evaluated and the vector of true breeding values ($\mathbf{TBV}$), $r(\mathbf{TBV}, \mathbf{EBV})$, cannot be computed. In the literature, multiple criteria have been proposed to quantify and compare prediction accuracies of genomic predictions between training and validation set structures and between prediction methods. Cross-validation approaches are often conducted based on $r(\mathbf{EBV}, \mathbf{y}^*)$, where $\mathbf{y}^*$ is either the vector of phenotypes adjusted for fixed effects or the vector of deregressed EBV of the validation set. Thus, a widely used criterion is $r(\mathbf{EBV}, \mathbf{y}^*)/\sqrt{h^2}$, where h^2 is the heritability of the trait. However, this criterion requires all the genotyped animals to have a sufficiently accurate $\mathbf{y}^*$ value (Legarra & Reverter, 2018). When $\mathbf{y}^*$ is an adjusted phenotype of the animal's own measurement, it suffers from the inability to adjust for the random residual effects. In the optimum situation, the expected value of the correlation would then be the square root of heritability (Gunia et al., 2014). Alternatively, using an EBV obtained from a complete dataset as the best predictor of TBV would cause autocorrelation between the reference and evaluated EBV when the training and validation sets are closely related through the pedigree, leading to greater correlations (Gunia et al., 2014). Legarra and Reverter (2018) proposed to complement the cross-validation approach with a semi-parametric approach that can be used in a large number of cases, with the advantage of not requiring knowledge of the TBV or adjustment of phenotypes. The underlying assumptions of this approach are (1) the variance components are similar in the training and validation datasets, and (2) the validation set is sufficiently diverse and large (i.e. composed of various families). In brief, with their approach, the correlation between EBV using part of the dataset (partial) and EBV obtained using the whole dataset results in an estimator of the ratio of the accuracies of the EBV from these two datasets. We followed this approach to evaluate the potential for genomic prediction when including data from a related line compared to genomic prediction using all data from the target line, which will be referred to as $GEBV_w$ (GEBV obtained using the whole dataset). I.e., to obtain $GEBV_w$ for the validation set of each line, two separate ssGBLUP analyses were performed (one per line). $GEBV_p$ (GEBV obtained using partial dataset) were the GEBV obtained from the six scenarios for the validation

sets in each target line. The criterion for prediction accuracy for each trait and each scenario was then the correlation between $\mathrm{GEBV_p}$ and $\mathrm{GEBV_w}$, $\mathrm{r(GEBV_p, GEBV_w)}$. Bias of the genomic predictions was computed as the deviation of the regression coefficient of $\mathrm{GEBV_w}$ on $\mathrm{GEBV_p}$ from 1, as also proposed in (Legarra & Reverter, 2018). Standard errors of the prediction accuracy correlations, *r*, were obtained as $\sqrt{[(1-\mathrm{r}^2)/(n-2)]}$, where n is the number of animals used to obtain correlations in the validation sets. Differences between correlations in different scenarios were tested using the Williams t-test in the psych R package (Revelle, 2019; Steiger, 1980; Williams, 1959). Significant differences between each scenario and the control scenario (scenario 1) are reported to identify the scenarios that provide prediction accuracies similar to the control scenario.

2.2.3.7. Relationships between training and validation sets

For each scenario, the maximum, average, and minimum relationship coefficients between training and validation sets in the $\mathbf{H}$ matrix were computed. To distinguish the strength of relationships originating from the two lines, all three measurements were computed separately for pigs of the validation set with the subset of the training set that belonged to 1) the target line and 2) the other line. The average relationships were calculated as the mean of the off-diagonal elements of the corresponding relationship matrices for the genotyped individuals.

2.2.4. Results

2.2.4.1. Variance components

The five studied traits showed low to moderate heritabilities that ranged from 0.12 ± 0.02 (RFI) to 0.36 ± 0.05 (BFT) (Table 2-4). The ratio of litter effect variance to phenotypic variance (l^2) was lower than the heritability for all traits, ranging from 0.07 ± 0.02 (FCR) to 0.12 ± 0.02 (BFT).

Table 2-4. Estimates of variance components (SE) of the studied traits

Trait	Phenotypic variance	Heritability	Litter effects[a]
ADG	5811.70 (164.75)	0.25 (0.04)	0.10 (0.02)
BFT	14.37 (0.47)	0.36 (0.05)	0.12 (0.02)
DFI	0.04 (0.001)	0.24 (0.04)	0.09 (0.02)
FCR	0.04 (0.001)	0.24 (0.04)	0.07 (0.02)
RFI	0.01 (0.004)	0.12 (0.02)	0.08 (0.02)

ADG average daily gain (g/day), *BFT* backfat thickness (mm), *DFI* daily feed intake (kg/day), *FCR* feed conversion ratio (kg/kg), *RFI* residual feed intake (kg/day), [a] As a proportion of phenotypic variance

2.2.4.2. Prediction accuracies

Prediction accuracies, $\mathrm{r}(\mathrm{GEBV_p}, \mathrm{GEBV_w})$, for the different scenarios are shown in Figure 2-2 for the two lines. Accuracies ranged from 0.07 to 0.73, depending on the validation line, trait, and scenario. The tested scenarios could be classified into two groups based on their design and how it affected the prediction accuracy of each trait. Removing the earlier generations of the target line from the training set (from scenarios 1, 2, 3 to scenarios 4, 5, 6) tended to decrease the prediction accuracy for ADG, BFT, and DFI, while FCR and RFI showed different patterns in response to changes in the structure of the training set.

The differences in prediction accuracies for ADG, BFT and DFI from scenario 1 to scenario 2 and 3 showed that the inclusion of different generations of the other line in the training set led to marginal changes in accuracy, with decreased correlations in most cases (BFT in the HRFI line and DFI). In scenarios 4, 5, and 6, the proportion of animals from the target line was low in the training set compared to scenarios 1, 2, and 3. This reduction generally led to a decrease in the prediction accuracies for ADG, BFT, and DFI compared to scenario 1. However, these differences in accuracy were only significant for ADG and BFT in the HRFI line and for DFI in the LRFI line.

Scenarios for FCR and RFI showed different patterns compared to the previous traits. Prediction accuracies for FCR followed a pattern similar to those of the other traits for all scenarios, except for scenario 3, which showed a 17 to 21% greater accuracy

compared to scenario 1. Prediction accuracies for RFI decreased from scenario 1 to scenario 2, and scenario 1 to scenario 4 for the LRFI target line, which were the scenarios with the maximum number of individuals from the other line in the training set. In the other scenarios, the prediction accuracies for RFI were similar or higher than for scenario 1.

The prediction accuracies for FCR in all scenarios, except scenario 6, were higher for validation animals in the HRFI line than in the LRFI line. The average differences in accuracy by trait ranged from +0.07 for ADG to +0.40 for RFI. (Figure 2-2).

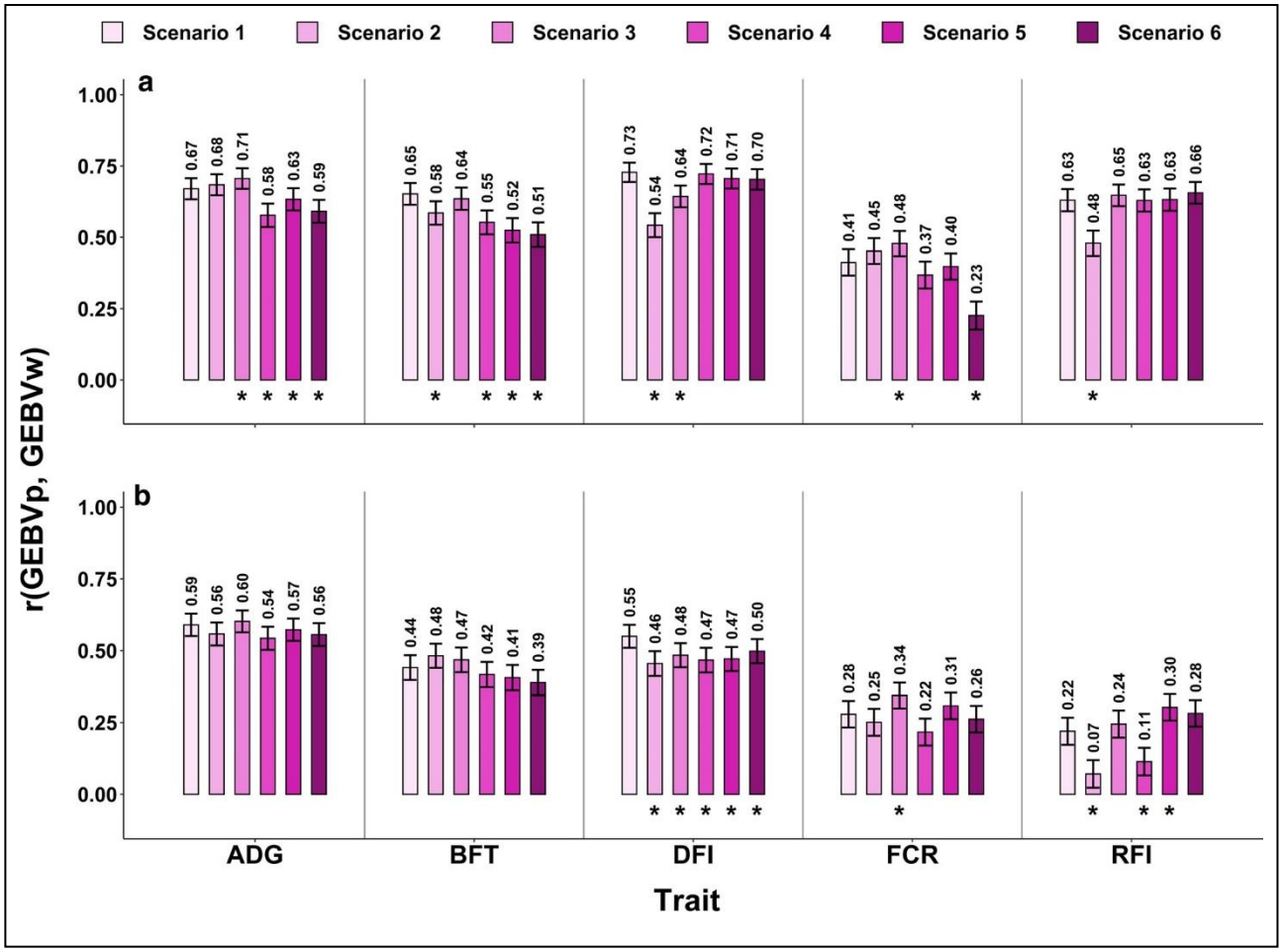


Figure 2-2. Correlations between GEBVp and GEBVw, and their SE as error bars for the HRFI (a) and LRFI (b) lines.

*Significant difference with scenarios 1 (control) based on the Williams t-test at a 0.05 level. RFI residual feed intake, ADG average daily gain, FCR feed conversion ratio, DFI daily feed intake, BFT backfat thickness

2.2.4.3. Prediction biases

Overall, regression coefficients of $\mathrm{GEBV_w}$ on $\mathrm{GEBV_p}$ were consistently below 1 for FCR and RFI for both validation sets (Figure 2-3). Regression coefficients for these two traits also showed more variation across the scenarios compared to ADG, BFT and DFI.

Bias for GEBV in the HRFI validation set followed the same trend, but at different magnitudes, for all traits, except ADG (Figure 2-3a). On average, scenarios 1, 2, and 3 showed less biases than scenario 4, 5, and 6 for BFT, DFI, and FCR. The regression coefficient in scenario 1 was equal to 0.98 for RFI, slightly over 1 for BFT (1.08) and DFI (1.19), and below 1 for ADG (0.83) and FCR (0.74).

Prediction of GEBV for the LRFI validation set did not follow the same pattern of change across scenarios between the traits. Regression coefficients of all scenarios showed biases lower than 1 for BFT, FCR, and RFI (Figure 2-3b). Biases were smallest for DFI (scenario 6) and ADG (scenarios 1, 5 and 6). Overall, biases of GEBV for this line were moderate for scenario 6 compared to the other scenarios, except for BFT (0.53). Biases were larger for scenarios 2 and 4, compared to scenarios 5 and 6, for all traits except for BFT.

2.2.4.4. Relationships between and within training and validation sets

Relationships between the validation set and the training individuals from the target line were on average higher in scenarios 4 to 6 than in scenarios 1 to 3 (Figs. 4a and 4c). The highest average was obtained for scenario 4 (around 0.25) and the smallest average for scenarios 1 and 3 (around 0.16 and 0.17). The maximum relationship coefficient between these two cohorts was greater than 0.66 for all scenarios, with the smallest maximum found for scenario 1 when the training set included only individuals from the target line, and the highest maximum for scenario 4 (around 0.78), when the relative number of animals from the other line in the training set was larger.

Relationship coefficients between the validation set and the training individuals of the other line were lower than those with the training individuals of the target line, but the

maximum values were reached for scenario 6, i.e. equal to 0.18 and 0.20 for the HRFI and LRFI target lines, respectively (Figs. 4b and 4d). All other scenarios had lower maximum relationships, ranging from 0.12 to 0.15.

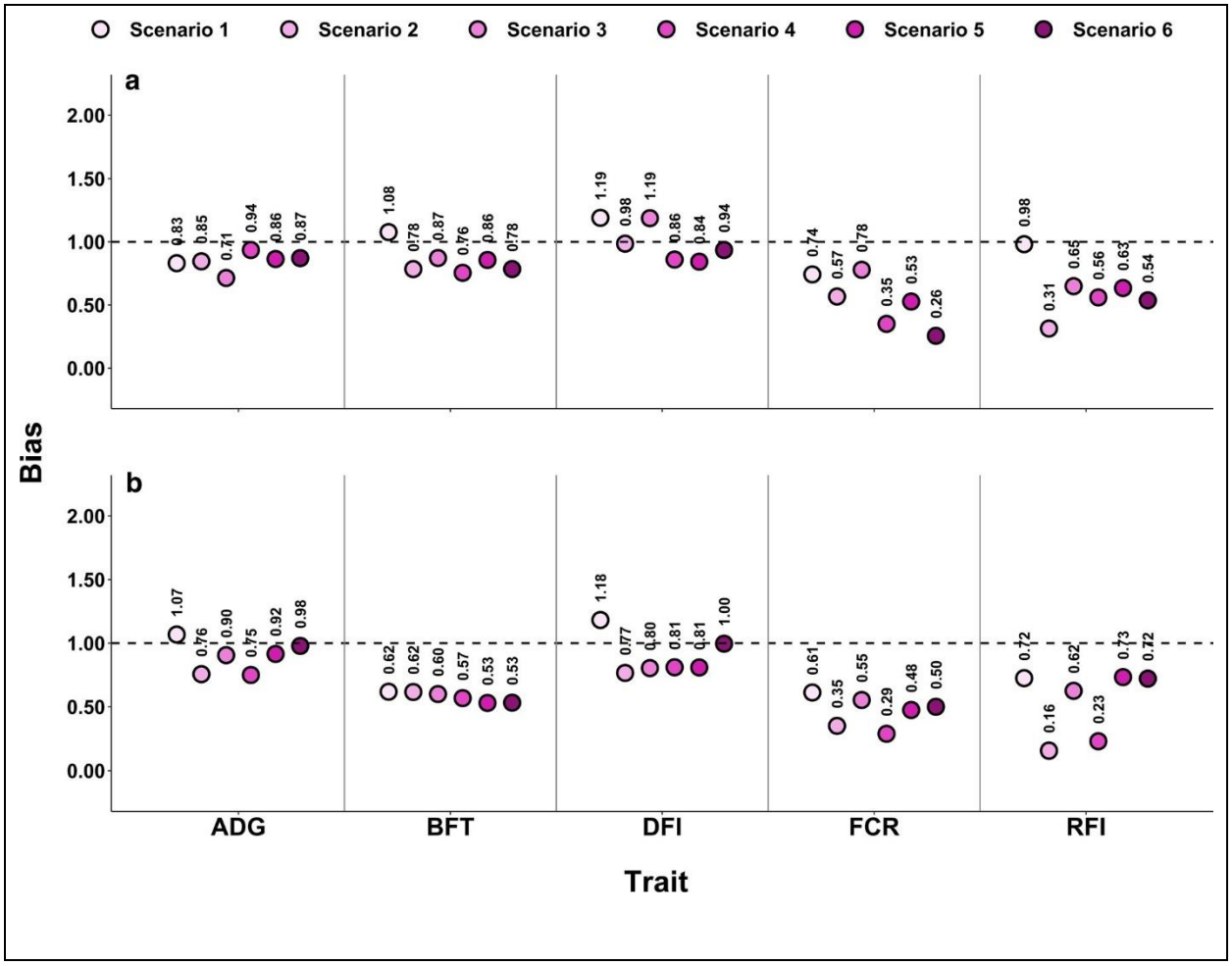


Figure 2-3. Bias (regression coefficients of GEBVw on GEBVp) for the HRFI (**a**) and LRFI (**b**) lines.

RFI residual feed intake, *ADG* average daily gain, *FCR* feed conversion ratio, *DFI* daily feed intake, *BFT* backfat thickness

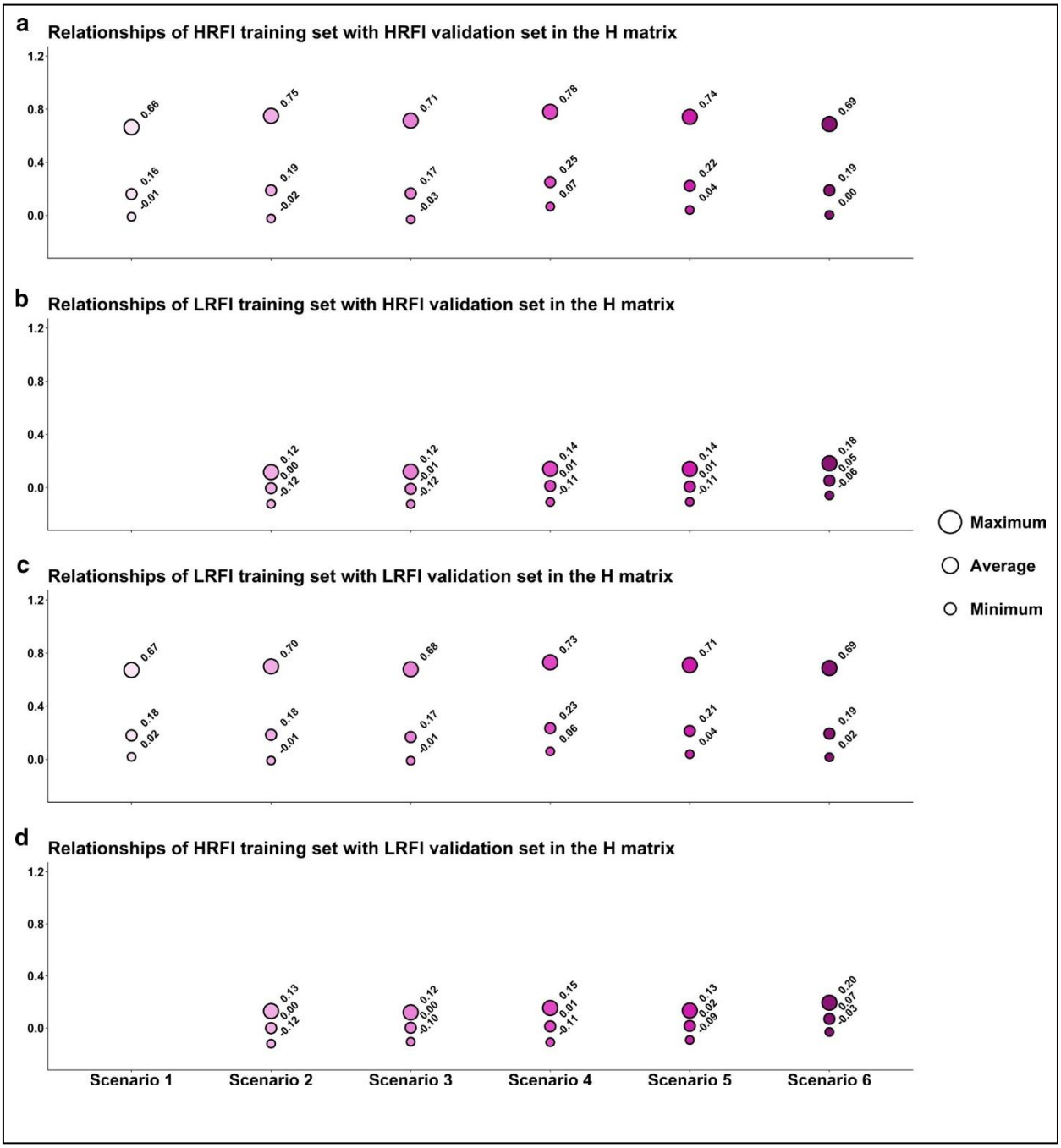


Figure 2-4. Average, minimum and maximum relationship coefficients in the H matrix between individuals of the validation set, and individuals of the training set from the target line and from the reverse line, for **a** and **b** the HRFI target line, for **c** and **d** the LRFI target line

2.2.5. Discussion

The aim of our study was to investigate different combinations of two lines derived from a common origin to evaluate the potential of building a training set for the genomic prediction of feed efficiency related traits in lines that are small or do not have much data available. Multiplying by ~ 2.5 (scenario 2), ~ 2 (scenarios 3 and 4), and ~ 1.5 times (scenario 5) the number of genotyped individuals in the training set by recruiting animals from the other line show no or little increase of prediction accuracy. This would probably not justify the additional genotyping costs involved. However, they can be considered for practical implementation of combined training sets since, in most cases, the prediction accuracies obtained in scenarios 5 and 6 were similar to those of the control scenario 1. These scenarios reflect most of the practical situations targeted in our study. Indeed, for breeding programs in small populations, phenotypic or genotypic information of individuals from earlier generations might not be available, and the sampling size in recent generations might be limited to a few hundred. Our results show that, a training population that includes recent generations of one population and data from a more distant subpopulation, could be a solution to achieve prediction accuracies similar to what would be achieved if data were available for individuals of the same population. This could even improve the prediction accuracies for traits under selection.

2.2.5.1. Computation of prediction accuracies and biases

Variance components of the evaluated traits were estimated using the **A** matrix on the full dataset with both lines combined. All estimated heritabilities were in the range of values reported in the literature for these traits (Christensen et al., 2012; de Campos et al., 2015; Do et al., 2014; Guo et al., 2016; Jiao et al., 2014). Using these variance components, the accuracy of GEBV was computed for the six scenarios to predict validation animals from each line using ssGBLUP. Prediction accuracies were computed using a cross-validation method combined with a semi-parametric approach [30]. Indeed, in our case, accuracies of the adjusted phenotypes or of deregressed EBV were too low to be used in a criterion such as $\mathrm{r}(\mathrm{GEBV}_{\mathrm{p}}, \mathrm{y}^{*})/\sqrt{\mathrm{h}^{2}}$, since only two thirds of the individuals had their own phenotype. This would result in larger standard errors of the correlations and, thus, less

power to test differences between scenarios, as shown in Figures S1 and S2 [see Additional file 1 Figures S1 and S2]. The underlying assumptions of the semi-parametric approach are that (1) the validation set is sufficiently diverse and large (i.e. composed of various families), and (2) variance components are similar in the training and validation datasets. The first assumption was well covered in our study, since all breeding individuals, plus some progeny of each family, were phenotyped and genotyped. The second assumption was potentially less covered, which could explain some of the biases in prediction observed. Indeed, when estimating variance components separately in the two lines, different residual variances were estimated for some traits, resulting in lower heritability estimates for DFI (24%), FCR (43%), and RFI (22%) in the LRFI line than in the HRFI line. Legarra and Reverter [30] indicated that inflation of predictions in one or the other dataset due to changes in variances can cause biased GEBV. Thus, we also tested the use of estimates of variance components from the target line for the GEBV predictions, but this resulted in increases in biases by 0.016 to 0.121 in all situations but one (results not shown). In practice, scaling the observations by the residual or phenotypic standard deviations, or accounting for the heterogeneity of residual variance across lines, could be considered to account for such differences, as proposed by Reverter et al. (1997) for heterogeneous variances across herds. An alternative could be to run bivariate analyses to consider correlated traits in the two lines, instead of a single trait across the two lines. Nevertheless, in our populations, estimates of the genetic variance of RFI as the trait under selection were consistent over the nine generations in each line. Therefore, differences in observed accuracy and bias between lines could not be explained by the heterogeneity of the genetic variance over the nine generations for the trait under selection.

2.2.5.2. Prediction accuracies for production traits

Although production traits and ssGBLUP have been discussed in the literature, few investigations have analyzed such traits in pigs with this method. Therefore, in the discussion that follows, we refer to published genomic prediction studies on these traits that often use other methods. Our objective in this part is to validate the prediction accuracies obtained with scenario 1, in which the structure of the training population is

close to those of previous studies. When comparing studies, it is worth noting that ssGBLUP generally has a higher accuracy than the usual GBLUP or Bayesian approaches that use only data of genotyped animals. Thus in theory, the comparisons should favor ssGBLUP approaches. However, most previous studies were based on prediction to a single generation of candidates, which could favor higher prediction accuracies. Despite these differences, overall, our estimates were within the range of accuracies reported in the literature, except for FCR and RFI, for which accuracies were higher in the HRFI validation set and lower in the LRFI line than those reported in the literature. In an investigation on 8113 Danish Duroc pigs with 60K imputed SNP genotyping information, an $\mathrm{r}(\mathrm{GEBV_p}, \mathrm{y}^*)/\sqrt{\mathrm{h}^2}$ of 0.41 was reported for ADG (Guo et al., 2016). In a study with 620 commercial boars, an $\mathrm{r}(\mathrm{GEBV_p}, \mathrm{y}^*)/\sqrt{\mathrm{h}^2}$ of 0.61 was reported for BFT with ridge regression BLUP (RR-BLUP) and of 0.56 with Bayesian LASSO (de Campos et al., 2015). A similar value of 0.55 was reported for Danish Duroc pigs (Guo et al., 2016). Zhang et al. (C. Y. Zhang et al., 2018) reported an $\mathrm{r}(\mathrm{GEBV_p}, \mathrm{y}^*)/\sqrt{\mathrm{h}^2}$ of 0.38 for DFI in a Duroc population using a 80K SNP chip and the GBLUP method in a design with 1167 training animals and 196 validation animals. They reported a higher accuracy (0.45) when using a 650k SNP chip and the BayesB method. Prediction accuracies of GEBV for FCR and RFI are rarely reported in the literature. Christensen et al. (2012) reported a prediction accuracy of 0.16 for FCR using a bivariate ssGBLUP model. Jiao et al. (2014) obtained a low prediction accuracy of 0.09 for RFI (measured as $\mathrm{r}(\mathrm{GEBV_p}, \mathrm{y}^*)/\sqrt{\mathrm{h}^2}$) using the BayesA method with 1047 training animals and 516 validation animals for the Duroc boars. Altogether, in pig studies, prediction accuracies are thus low to moderate for ADG and BFT, and low for feed efficiency traits.

2.2.5.3. Prediction accuracies depending on the training set composition

Compared to FCR and RFI, ADG, BFT, and DFI showed different prediction accuracy changes compared to the scenario 1 when the structure of the training set was changed. For ADG, BFT, and DFI, removing the earlier generations of the target line from the training set (from scenarios 2 and 3 to scenarios 4, 5 and 6) generally decreased prediction accuracy to a lesser extent. The average and maximum relationships between

the validation set and the training subsets were higher in scenarios 4, 5, and 6 than in scenario 1. The maximum relationship between the validation set and the training subsets, previously recommended as an indicator of potential accuracies (Clark et al., 2012), was lowest in scenario 1 and highest in scenario 4, likely due to changes in allele frequencies between the early and late generations within a line. This implies that the general decrease in accuracy in the scenarios 4, 5, and 6 could neither be attributed to these changes in relationships between sets, nor to the differences in prediction accuracies between lines. Moreover, the accuracy of GEBV resulting from ssGBLUP analyses should be less sensitive to the structure of the set of genotyped animals, and accordingly, to the strength of relationships between and within training and validation sets (Lourenco et al., 2015) because the $\mathbf{H}$ matrix aggregates information from both $\mathbf{A}$ and $\mathbf{A_{22}}$. This structure of the **H** matrix has two major effects on the GEBV of a given animal: first, it contributes the parent average EBV of the animal using the $\mathbf{A}$ matrix, and second, it adjusts for the different levels of relationships of the animal with other genotyped animals using the $\mathbf{A_{22}}$ matrix (Lourenco et al., 2015; Misztal et al., 2013). de Roos et al. (2009) reported that the benefits of combining populations in a training set are greatest when the populations have diverged for only a few generations and when the heritability of the trait is low. They also showed that increasing the number of animals from a given population in the training set increased prediction accuracy in that population. Considering that de Roos et al. (2009) did not include the effect of selection in their simulations, this could partly explain our results for ADG, BFT, and DFI.

2.2.5.4. Impact of selection on accuracy and bias of predictions

The changes of accuracy across the scenarios were more diverse for RFI and FCR, with either increases or relatively similar accuracies compared to scenario 1. In some cases, the accuracy even increased as genotypes of closer generations were eliminated from the training set, which could be regarded as an effect of the different relationships between training and validation sets in these scenarios. Regarding the low prediction accuracy reported for FCR and RFI in our results and in the literature, denser SNP genotyping could probably increase the accuracy of predictions by better capturing the differences in LD between the lines. In addition, for low heritability traits, such as RFI in

our study, large training populations have been reported to increase the accuracy of GEBV (Goddard, 2009; B J Hayes et al., 2009; Hoze et al., 2014). However, given that scenarios 5 and 6 resulted in accuracies that were comparable to that of the control scenario for FCR and in greater accuracies for RFI, they can be considered as optimum scenarios for an across-line genomic prediction program. Based on results from simulation, Pszczola et al. (2012) declared that minimizing relationships within the reference population and maximizing them between training and validation sets maximizes the accuracy of genomic predictions. This means that including a diverse set of animals in the training set is desirable to some extent. This is consistent with our results for FCR and RFI, for which selection created two diverse sets of animals. For example, in scenario 6, including animals from G4 to G6 of the target line in the training set provided sufficient genetic links between training and validation sets, and animals from the G9 generation of the other line provided additional diversity to the training set. Overall, it seems that including animals from later generations of both lines (more diverse animals) in the training set contributed to greater accuracies of GEBV in the validation set for FCR and RFI. This might be because the SNP effects segregating in the validation set were better estimated with such a training set.

Overall, the comparison of accuracies between scenarios 4 to 6 and scenario 1 did not show an obvious effect of the removal of data of earlier generations from the training dataset. In a study using six levels of truncated data of past generations, accuracies of GEBV of young genotyped pigs were very similar for various reproductive traits (Lourenco et al., 2014).

2.2.5.5. Bias of genomic predictions

Our results showed that GEBV were more biased for traits that were more affected by selection, especially when early generations of the target line were not included in the training set. The scenarios that yielded better accuracies were not those with the smaller biases, except for FCR and RFI, for which predictions were low and their regression coefficients were systematically below 1. The average and maximum relationships between training and validation sets did not affect the prediction biases in the same way

for all traits, which could be due to the effect of selection. Selection in historical generations has been shown to result in considerable biases in EBV or GEBV (Bijma, 2012; Legarra & Reverter, 2018). Tonussi et al. (2017) emphasized that, to have accurate and unbiased GEBV with the ssGBLUP method, the $\mathbf{G}$ matrix should be compatible with the $\mathbf{A_{22}}$. Inappropriate merging of these matrices can originate from ignoring inbreeding in the structure of $\mathbf{A}$ and from changes in allele frequencies at QTL for the traits under selection. In our scenarios, the effect of selection in the last three generations of the validation sets was not explicitly accounted for. However, changes in marker allele frequencies in those generations were accounted for through the $\mathbf{G}$ matrix. Furthermore, the (co)variances used for genomic predictions were obtained from bivariate analyses including selection criterion using the whole dataset (including validation generations). Therefore, there should be no effect of selection on the estimations of the variance components, and the prediction bias of the GEBVs should not be due to biased variance components. Computing separate accuracies and biases for sires (heavily selected) versus dams (not directly selected), could enable quantifying the effect of selection on the prediction biases. However, on the one hand, the dams had lower individual accuracies (no own phenotype), and on the other hand, only 18 sires were selected per line in these generations. Therefore, the resulting prediction accuracies and biases differed between sires and dams due to factors other than just the effect of selection and no clear conclusion could be reached. Finally, it should be mentioned that these three generations were combined into the validation set in our study to have enough individuals, but in practice, new candidates to be predicted pertain to a single unselected cohort, therefore this selection effect would be low and likely negligible.

Heritability, marker density and size of the training population have been shown to be important factors to control biases of prediction (Karimi et al., 2019). Therefore, the biases for some scenarios in this study could be explained by the low to medium heritability of the traits, the medium marker density information, and the low number of individuals in the training population. Testing similar prediction scenarios while ignoring pedigree relationships in the non-genotyped generations would lead to substantially biased predictions, especially for traits affected by selection (for instance, 1.61 for RFI

predictions in the HRFI line for scenario 6). Combining full pedigree and genomic information appeared to limit bias, which is consistent with Tonussi et al. (2017).

2.2.6. Conclusions

The results of our study show that genomic prediction using a training set that includes animals from related lines selected in different directions could be as accurate as genomic prediction using a within-line training set. Thus, this can be a solution to create a reference set in the case of small populations, or when ancestral samples are not available at low additional costs. Combined reference sets had better prediction accuracies for traits that were highly affected by selection, which can be attributed to the inclusion of more diverse animals in the training set. Overall, among all evaluated scenarios, scenarios 5 and 6 showed optimal accuracies in most cases, consistent with our hypothesis that data from a related line can be used in a combined training population for genomic predictions without losing prediction accuracy. Our results also proved that absence of phenotypic records from past generations did not affect prediction accuracy but increased bias of predictions. Some of these issues could be solved by using bivariate analyses or models with heterogeneous variances to better account for changes in variances with selection in different lines. Altogether, the results of our study provide insights into the design of reference populations for small populations, particularly when lines are being developed simultaneously, which is common in poultry and pig industries, as well as some plant breeding plans. This strategy can be recommended to initiate a genomic selection program when historical samples are not available, or when two lines are considered and genotyping costs need to be limited.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AA performed the statistical analyses and wrote the first draft of the paper. ED and YL performed the imputation and quality control of the genotypic data. AA, JR and HG

participated in the design of the study. HG provided scientific supervision. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Andres Legarra for his comments and help. This study was financially supported by the French National Research Agency via the MicroFeed project, under grant ANR-16-CE20-0003. The authors would like to thank the experimental farm staff for data collection and breeding of the animals. The authors are also grateful to Yvette Steyn for proof-reading the manuscript.

2.2.7. References

Aguilar, I., Misztal, I., Johnson, D., Legarra, A., Tsuruta, S., & Lawlor, T. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, *93*(2), 743-752.

Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A., Resende Jr, M. F., & Muñoz, P. R. (2016). AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. *The Plant Genome*, *9*(3), 1-10.

Bijma, P. (2012). Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *Journal of Animal Breeding and Genetics*, *129*(5), 345-358.

Careau, V., Wolak, M. E., Carter, P. A., & Garland Jr, T. (2013). Limits to behavioral evolution: the quantitative genetics of a complex trait under directional selection. *Evolution*, *67*(11), 3102-3119.

Carillier, C., Larroque, H., & Robert-Granie, C. (2014). Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. *Genetics Selection Evolution*, *46*, 67. https://doi.org/10.1186/s12711-014-0067-3

Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, *42*(1), 2.

Christensen, O. F., Madsen, P., Nielsen, B., Ostersen, T., & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal*, *6*(10), 1565-1571. https://doi.org/10.1017/S1751731112000742

Clark, S. A., Hickey, J. M., Daetwyler, H. D., & van der Werf, J. H. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genetics Selection Evolution*, *44*(1), 4.

de Campos, C. F., Lopes, M. S., e Silva, F. F., Veroneze, R., Knol, E. F., Lopes, P. S., & Guimarães, S. E. (2015). Genomic selection for boar taint compounds and carcass traits in a commercial pig population. *Livestock Science*, *174*, 10-17.

de Roos, A. P., Hayes, B. J., & Goddard, M. E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, *183*(4), 1545-1553. https://doi.org/10.1534/genetics.109.104935

Do, D. N., Janss, L. L., Strathe, A. B., Jensen, J., & Kadarmideen, H. (2014). Genomic prediction and genomic variance partitioning of daily and residual feed intake in pigs using Bayesian Power Lasso models. The 10th World Congress on Genetics Applied to Livestock Production,

Fangmann, A., Bergfelder-Drüing, S., Tholen, E., Simianer, H., & Erbe, M. (2015). Can multi-subpopulation reference sets improve the genomic predictive ability for pigs? *Journal of Animal Science*, *93*(12), 5618-5630.

Fernando, R., & Gianola, D. (1990). Statistical inferences in populations undergoing selection or non-random mating. In *Advances in statistical methods for genetic improvement of livestock* (pp. 437-453). Springer.

Gaines, A., Peterson, B., & Mendoza, O. (2012). Herd management factors that influence whole herd feed efficiency. In *Feed efficiency in swine* (pp. 15-39). Springer.

Gilbert, H., Bidanel, J. P., Gruand, J., Caritez, J. C., Billon, Y., Guillouet, P., Lagant, H., Noblet, J., & Sellier, P. (2007). Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. *Journal of Animal Science*, *85*(12), 3182-3188. https://doi.org/10.2527/jas.2006-590

Gilbert, H., Billon, Y., Brossard, L., Faure, J., Gatellier, P., Gondret, F., Labussiere, E., Lebret, B., Lefaucheur, L., Le Floch, N., Louveau, I., Merlot, E., Meunier-Salaun, M. C., Montagne, L., Mormede, P., Renaudeau, D., Riquet, J., Rogel-Gaillard, C., van Milgen, J., Vincent, A., & Noblet, J. (2017). Review: divergent selection for residual feed intake in the growing pig. *Animal*, *11*(9), 1427-1439. https://doi.org/10.1017/S175173111600286X

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, *136*(2), 245-257. https://doi.org/10.1007/s10709-008-9308-0

Gunia, M., Saintilan, R., Venot, E., Hoze, C., Fouilloux, M. N., & Phocas, F. (2014). Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *Journal of Animal Science*, *92*(8), 3258-3269.

Guo, X., Christensen, O. F., Ostersen, T., Wang, Y., Lund, M. S., & Su, G. (2016). Genomic prediction using models with dominance and imprinting effects for backfat thickness and average daily gain in Danish Duroc pigs. *Genetics Selection Evolution*, *48*(1), 67.

Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., & Kruuk, L. E. (2010). The misuse of BLUP in ecology and evolution. *The American Naturalist*, *175*(1), 116-125.

Hayes, B. J., Bowman, P. J., Chamberlain, A., & Goddard, M. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, *92*(2), 433-443.

Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K., & Goddard, M. E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, *41*(1), 51.

Henderson, C. (1990). Accounting for selection and mating biases in genetic evaluations. In *Advances in statistical methods for genetic improvement of livestock* (pp. 413-436). Springer.

Hoque, M., Suzuki, K., Kadowaki, H., Shibata, T., & Oikawa, T. (2007). Genetic parameters for feed efficiency traits and their relationships with growth and carcass traits in Duroc pigs. *Journal of Animal Breeding and Genetics*, *124*(3), 108-116.

Hoze, C., Fritz, S., Phocas, F., Boichard, D., Ducrocq, V., & Croiseau, P. (2014). Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. *Journal of Dairy Science*, *97*(6), 3918-3929.

Jiao, S., Maltecca, C., Gray, K., & Cassady, J. (2014). Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: I. Genetic parameter estimation and accuracy of genomic prediction. *Journal of Animal Science*, *92*(6), 2377-2386.

Karimi, K., Sargolzaei, M., Plastow, G. S., Wang, Z., & Miar, Y. (2019). Opportunities for genomic selection in American mink: A simulation study. *PloS one*, *14*(3), e0213873.

Koch, R. M., Swiger, L. A., Chambers, D., & Gregory, K. E. (1963). Efficiency of Feed Use in Beef Cattle 1. *Journal of Animal Science*, *22*(2), 486-494.

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, *92*(9), 4656-4663. https://doi.org/10.3168/jds.2009-2061

Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution*, *50*(1), 53. https://doi.org/10.1186/s12711-018-0426-6

Lourenco, D., Misztal, I., Tsuruta, S., Aguilar, I., Lawlor, T., Forni, S., & Weller, J. (2014). Are evaluations on young genotyped animals benefiting from the past generations? *Journal of Dairy Science*, *97*(6), 3930-3942.

Lourenco, D. A., Fragomeni, B. O., Tsuruta, S., Aguilar, I., Zumbach, B., Hawken, R. J., Legarra, A., & Misztal, I. (2015). Accuracy of estimated breeding values with genomic information on males, females, or both: an example on broiler chicken. *Genetics Selection Evolution*, *47*(1), 56.

Lund, M. S., Su, G., Janss, L., Guldbrandtsen, B., & Brondurn, R. F. (2014). Invited review: Genomic evaluation of cattle in a multi-breed context. *Livestock Science*, *166*, 101-110. https://doi.org/10.1016/j.livsci.2014.05.008

Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, *92*(9), 4648-4655. https://doi.org/10.3168/jds.2009-2064

Misztal, I., Tsuruta, S., Aguilar, I., Legarra, A., VanRaden, P., & Lawlor, T. (2013). Methods to approximate reliabilities in single-step genomic evaluation. *Journal of Dairy Science*, *96*(1), 647-654.

[Record #60 is using a reference type undefined in this output style.]

Ollivier, L., Gueblez, R., Webb, A., & Van der Steen, H. (1990). Breeding goals for nationally and internationally operating pig breeding organisations. Proceedings of the 4th World Congress on Genetics applied to Livestock Production, Edinburgh 23-27 July 1990. XV. Beef cattle, sheep and pig genetics and breeding, fibre, fur and meat quality.,

Olson, K., VanRaden, P., & Tooker, M. (2012). Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science*, *95*(9), 5378-5383.

Patience, J. F., Rossoni-Serão, M. C., & Gutiérrez, N. A. (2015). A review of feed efficiency in swine: biology and application. *Journal of animal science and biotechnology*, *6*(1), 33.

Pszczola, M., Strabel, T., Mulder, H., & Calus, M. (2012). Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science*, *95*(1), 389-400.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, *81*(3), 559-575.

Pym, R., & Nicholls, P. (1979). Selection for food conversion in broilers: Direct and correlated responses to selection for body-weight gain, food consumption and food conversion ratio. *British Poultry Science*, *20*(1), 73-86.

Revelle, W. (2019). Package Psych V1. 8.12: Procedures for Psychological, Psychometric, and Personality Research. *Evanston, IL: Northwestern University*.

Reverter, A., Tier, B., Johnston, D., & Graser, H.-U. (1997). Assessing the efficiency of multiplicative mixed model equations to account for heterogeneous variance across herds in carcass scan traits from beef cattle. *Journal of Animal Science*, *75*(6), 1477-1485.

Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, *15*(1), 478.

Sorensen, D., Fernando, R., & Gianola, D. (2001). Inferring the trajectory of genetic variance in the course of artificial selection. *Genetics Research*, *77*(1), 83-94.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological bulletin*, *87*(2), 245.

Tonussi, R. L., de Oliveira Silva, R. M., Magalhães, A. F. B., Espigolan, R., Peripolli, E., Olivieri, B. F., Feitosa, F. L. B., Lemos, M. V. A., Berton, M. P., & Chiaia, H. L. J. (2017). Application of single step genomic BLUP under different uncertain paternity scenarios using simulated data. *PloS one*, *12*(9), e0181752.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414-4423.

VanRaden, P. M., O'Connell, J. R., Wiggans, G. R., & Weigel, K. A. (2011). Genomic evaluations with many more genotypes. *Genetics Selection Evolution*, *43*, 10. https://doi.org/10.1186/1297-9686-43-10

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *21*(2), 396-399.

Zhang, C. Y., Kemp, R. A., Stothard, P., Wang, Z. Q., Boddicker, N., Krivushin, K., Dekkers, J., & Plastow, G. (2018). Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. *Genetics Selection Evolution*, *50*(1), 14. https://doi.org/ARTN 14

10.1186/s12711-018-0387-9

Zhang, S.-Y., Olasege, B. S., Liu, D.-Y., Wang, Q.-S., Pan, Y.-C., & Ma, P.-P. (2018). The genetic connectedness calculated from genomic information and its effect on the accuracy of genomic prediction. *PloS one*, *13*(7), e0201400.

2.2.8. Appendix 2.1

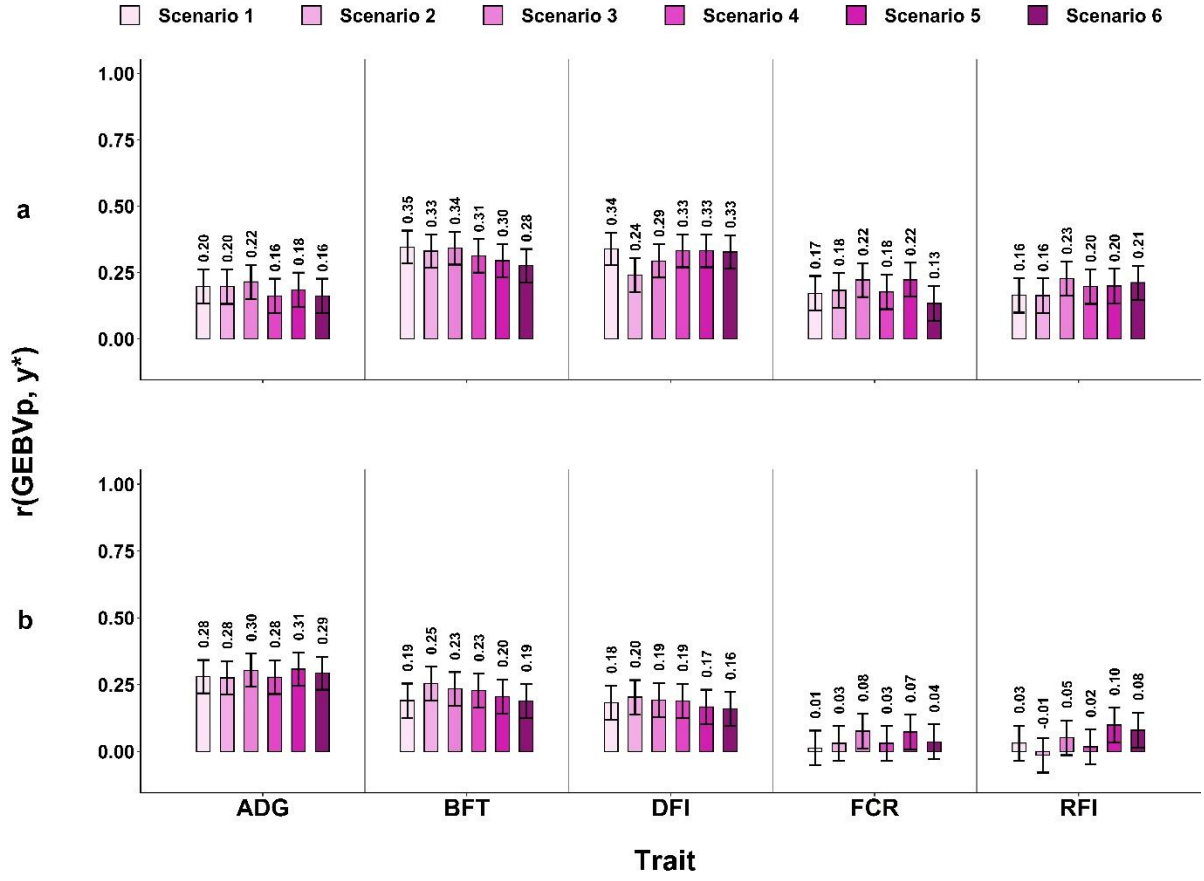


Figure S1. Correlation between $\mathrm{GEBV_p}$ and $\mathrm{y^*}$ and their SE as bars for the HRFI (a) and LRFI (b) lines. No scenario resulted in correlations that differed from those with scenario 1 based on a Williams t-test at 5%. RFI residual feed intake, ADG average daily gain, FCR feed conversion ratio, DFI daily feed intake, BFT backfat thickness.

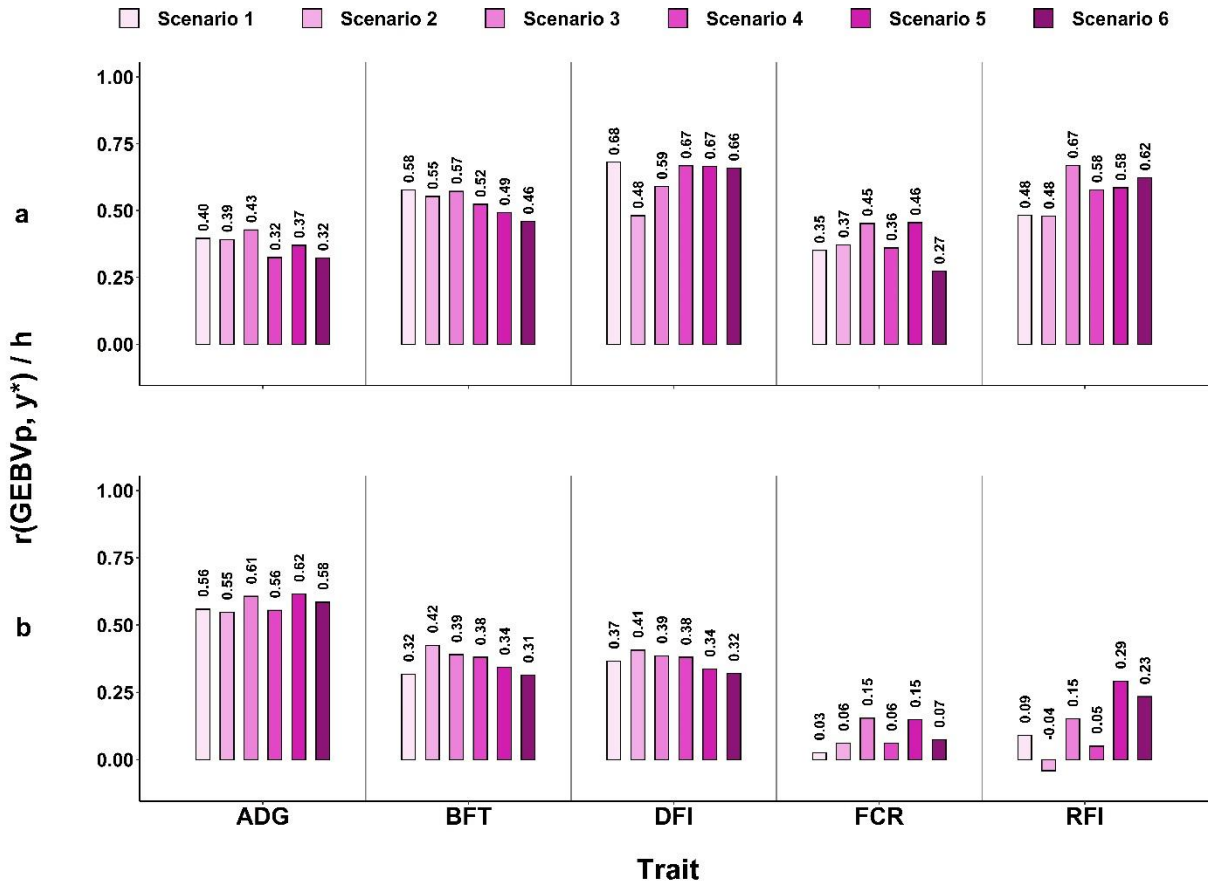


Figure S2. Correlation between $\mathrm{GEBV_p}$ and $\mathrm{y^*}$ divided by the square root of the heritability of corresponding traits for the HRFI (a) and LRFI (b) lines.

RFI residual feed intake, ADG average daily gain, FCR feed conversion ratio, DFI daily feed intake, BFT backfat thickness.

2.2.9. Appendix 2.2

Gordon Research conference: Quantitative genetics and genomics, Feb 2019, Barga, Italy. 2019

Reliability of the genomic predictions for the feed efficiency related trait based on different pig lines

Amir Aliakbari, Emilie Delpuech, Yann Labrune, Juliette Riquet, and Hélène Gilbert

GenPhySE, INRA, 31320 Castanet-Tolosan, FRANCE

The majority of genomic evaluations of feed efficiency have been studied on a unique population, divided in a reference and a validation set. However, genomic evaluations using genetically different reference and validation sets like divergently developed lines can provide a wider insight into the flexibility in size and structure of the reference population for genomic evaluations of feed efficiency in limited size swine populations. Therefore, the aim of our study was to investigate the possibility of genomic evaluation for feed efficiency related traits based on a reference population that comprised different combinations of animals from two different lines. The two pig lines have been established during 9 generations of divergent selection for the residual feed intake (RFI). We evaluated the accuracy of the predicted genomic breeding values (GBVs) using a single-step genomic BLUP (ssGBLUP) method through the eight different scenarios. All scenarios had the same validation set (last generations of a given line), but they differed in structure of the reference population. We found that some combinations of generations from the two lines provided acceptable accuracies of the GBVs, which could be beneficial for breeding plans of swine lines of limited size that are based on the genomic selection.

2.2.10. Appendix 2.3

70. Annual meeting of the European Federation of Animal Science (EAAP), Aug 2019, Ghent, Belgium. 717p.

Reliability of genomic predictions for feed efficiency traits based on different pig lines

Amir Aliakbari, Emilie Delpuech, Yann Labrune, Juliette Riquet, and Hélène Gilbert

GenPhySE, INRA, 31320 Castanet-Tolosan, FRANCE

The majority of genomic predictions use a unique population split between a reference and a validation set. However, a genomic evaluation using genetically different reference and validation sets could provide more flexibility for the choice of reference sets in small populations. The aim of our study was to investigate the potential of genomic evaluation for feed efficiency related traits using a reference set that combines two different lines. Data came from two lines divergently selected for residual feed intake during 9 generations. Genomic breeding values (GBVs) of animals for five production traits were predicted using the single-step genomic BLUP method with six scenarios. All scenarios aimed to predict GBVs of pigs of the three last generations (~ 400 pigs, G7 to G9) in one or in the other line (validation line). To compare the scenarios prediction accuracy, a first scenario (control) had a reference set with animals from G1 to G6 (~ 400 pigs) of the validation line. In scenario 2, in addition to those of the control scenario, the reference set included about 600 pigs from G4 to G9 of the alternate line. Scenario 3 had ~ 800 pigs in the reference set, by excluding animals from G4 to G6 of the alternate line from the reference set compared to scenario 2. For the last three scenarios, fewer animals from the validation line were included in the reference set (~200 pigs from G4 to G6). In

scenario 4, G4 to G9 animals from the alternate line (~600 pigs, as in scenario 2) were included in the reference set. In scenario 5, only ~400 pigs from G7 to G9, and in scenario 6 ~200 pigs from G9, were used. In scenarios 2, 3 and 4, genotyping 400 to 600 additional individuals from the alternate line provided on average limited improvement the prediction accuracies for the five traits (<14%, except in 3 cases), and sometimes led to reduced accuracies. Scenarios 5 and 6 had similar accuracies as the control scenario, with less genotyping in scenario 6. It indicates that if samples from earlier generations are missing in a line, part of them can be replaced by recent samples from a related different line, giving more flexibility to design training populations in small lines.

Keywords: divergent lines, feed efficiency, genomic prediction, prediction accuracy, swine

3.

Genetic basis of the gut microbiota and their relationships with production trait

3.1. Introduction

Recent advances in obtaining microbiota information enable surveying the interplay between complex traits and the microbial community of the gastrointestinal tract (GIT). To initiate the evaluation of the potential of gut microbiota to selection for feed efficiency, this chapter aimed to evaluate the genetic relationship between faecal microbial composition and five feed efficiency and production traits. A total of 588 samples from two experimental lines were sequenced for the 16 rRNA hypervariable V3-V4 region. First, the microbial communities were compared between genetic lines and then genera abundances and two α-diversity indexes were analysed using bivariate and three-variate animal linear mixed models to estimate the heritability (h^2) of the microbiota traits, and their genetic correlations (r_g) with the phenotypic traits.

In the first step, a non-metric multidimensional scaling showed line differences between genera, with significantly different loadings of the genera along the second axis of the analysis. In addition, the α-diversity indexes were higher in the LRFI line than in the HRFI line. With the genetic analyses, the h^2 estimates of these α-diversity indexes were 0.19 ± 0.08 (Shannon) and 0.12 ± 0.06 (Simpson). Among the 75 genera kept in the analyses, 48 genera had a significant h^2 (> 0.125, threshold obtained by bootstrapping the abundances across individuals). The r_g of the α-diversities indexes with production traits were negative and significantly different from zero. Some genera belonging to the *Lachnospiraceae*, *Ruminococcaceae*, *Prevotellaceae*, *Lactobacillaceae*, *Streptococcaceae*, *Rikenellaceae* and *Desulfovibrionaceae* families had r_g significantly different from zero with the three of the studied traits, RFI, DFI and BFT, suggesting a stronger genetic link between gut microbiota components and these traits than with FCR and ADG. These results showed that the gut microbial community and α-diversity indicators are partly heritable and have genetic relationships with FE, that offer promising perspectives for selection for feed efficiency using gut microbiome composition in pigs.

This chapter was published as a journal paper in Journal of Animal Breeding and Genetics (DOI: 10.1111/jbg.12539). The supplementary material of this paper is provided in the Appendix 3.1. Preliminary analyses on OTU were presented in an oral presentation at EAAP-2020 (Appendix 3.2).

3.2. Article II: Genetic relationships between feed efficiency and gut microbiome in pig lines selected for residual feed intake

Amir Aliakbari[1], Olivier Zemb[1], Yvon Billon[2], Celine Barilly[1], Ingrid Ahn[1], Juliette Riquet[1] and Helene Gilbert[1*]

[1]GenPhySE, Universite de Toulouse, INRAE, Castanet-Tolosan, France

[2]GenESI, INRAE, Surgeres, France

Journal of Animal Breeding and Genetics (2021), *Early online*

https://doi.org/10.1111/jbg.12539

3.2.1. Abstract

This study aimed to evaluate the genetic relationship between faecal microbial composition and five feed efficiency and production traits, residual feed intake (RFI), feed conversion ratio (FCR), daily feed intake (DFI), average daily gain (ADG) and backfat thickness (BFT). A total of 588 samples from two experimental pig lines developed by divergent selection for RFI were sequenced for the 16 rRNA hypervariable V3-V4 region. The 75 genera with less than 20% zero values (97% of the counts) and two α-diversity indexes were analysed. Line comparison of the microbiota traits and estimations of heritability (h^2) and genetic correlations (r_g) were analysed. A non-metric multidimensional scaling showed line differences between genera. The α-diversity indexes were higher in the LRFI line than in the HRFI line ($P < 0.01$), with h^2 estimates of 0.19 ± 0.08 (Shannon) and 0.12 ± 0.06 (Simpson). Forty-eight genera had a significant h^2 (> 0.125). The r_g of the α-diversities indexes with production traits were negative. Some r_g of genera belonging to the *Lachnospiraceae*, *Ruminococcaceae*, *Prevotellaceae*, *Lactobacillaceae*, *Streptococcaceae*, *Rikenellaceae* and *Desulfovibrionaceae* families significantly differed from zero ($P < 0.05$) with FE traits, RFI (3), DFI (7), and BFT (11). These results suggest that a sizable part of the variability of the gut microbial community is under genetic control and has genetic relationships with FE, including diversity indicators. It offers promising perspectives for selection for feed efficiency using gut microbiome composition in pigs.

Keywords

Feed efficiency, genetic, gut microbiome, heritability, pigs

3.2.2. Introduction

Recent advances in bioinformatics and sequencing technologies have made it possible to obtain individual microbiome information for humans, animals, and plants. The fundamental role of gut microbiota in essential biological processes such as physiological aging in humans (Muscogiuri et al., 2019), methane emission in dairy cows, and nutrient digestion, absorption, and metabolism of pigs (Q. Niu et al., 2019) makes it a key field of research to counteract major physiological defaults such as obesity in human, and to improve quantitative production traits in livestock. In this regard, measuring the magnitude of genetic control on gut microbiota composition is fundamental to enlighten its potential use in animal selection programs. From a quantitative genetics perspective, estimating heritability (h^2) quantifies the magnitude of genetic control of a trait. Heritability is a population-specific parameter that estimates the proportion of additive genetic variance to the phenotypic variance of the trait. Besides the heritability, another essential genetic parameter is the additive genetic correlation (r_g). These two parameters are crucial to predict direct and correlated responses to selection, which are other parameters to evaluate if and how a trait would be affected by selection (Brenner et al., 2002).

In pig breeding, production and feed efficiency (FE) traits, because of their key economic and environmental importance, have a high impact on the sustainability of this industry (Ottosen et al., 2020). Therefore, research around FE cover a wide range of studies, from traditional statistical methods to recent advances in benefiting from biological data like metabolomics, including few with microbiome information (Maltecca et al., 2020). Several previous studies attempted to discover the link between host genetics, microbiota data, and feed efficiency (Bergamaschi, Maltecca, et al., 2020; Bergamaschi, Tiezzi, et al., 2020a; Camarinha-Silva et al., 2017; U. M. McCormack et al., 2017). A study on low and high residual feed intake (RFI) pigs showed a slight difference between the intestinal microbiota of two groups of animals chosen for their phenotypic RFI, and suggested a link between microbial community and FE at the phenotypic level (U. M. McCormack et al., 2017). However, direction of correlated responses between RFI and microbiota composition are still unknown. In the present study, we aimed to seek the genetic relationships between five production and FE traits and faecal microbial

composition, using data from two experimental pig lines developed by divergent selection for RFI. Statistical analyses were applied to microbiota genera, microbial diversity and performance traits to compare faecal microbiota composition between lines, and h^2 and r_g were obtained to describe the transmissible relationships between these traits and microbial traits.

3.2.3. Materials and Methods

3.2.3.1. Data structure

The data were collected from two experimental French Large White pig lines developed during 10 generations of divergent selection for RFI between 2000 to 2017 at INRAE (UE GenESI, Surgères, France, https://doi.org/10.15454/1.5572415481185847E12). The selection process and structure of the data from the two divergent lines has been described in Gilbert et al. (2017) and Aliakbari et al. (2020). Briefly, the G0 individuals were obtained from artificial insemination of 30 sows with 30 boars in generation F0. From the G0 litters, 116 boars were tested for RFI as candidates for selection. Among them 6 extreme low RFI (LRFI) and 6 extreme high RFI (HRFI) boars were selected to be the founders of each line. The selected founder boars were randomly mated with about 70 G0 gilts to initiate the two divergent lines. From generations G1 to G10, the same procedure was implemented within each line, with 96 tested boars per line to produce the next generation. There was no selection on the female side, and sows from both lines were distributed in two farms in equal proportions, which corresponds to two herds of birth for the tested pigs. After weaning, all pigs were gathered on the same farm for testing. In each generation, at least one additional parity was produced to evaluate the correlated responses to selection of growth, feed intake and efficiency and carcass composition traits on females and castrated males (response animals). Candidates to selection were tested from 35 to 95 kg of body weight (BW), whereas for response animals the test ran from 10 weeks of age until slaughter (105 kg BW until G5 and 115 kg BW afterward). Testing was organised in four pens per contemporary group (CG), and there were at least four CG tested per generation, systematically including both lines. Pigs were penned in groups of 12, per line, and sex when multiple sexes were tested. Pens were equipped with single-place electronic feeders ACEMA64 (ACEMO, France) to record individual feed

intake. A pelleted diet based on cereals and soya bean meal was available ad libitum, and contained 10 MJ net energy (NE)/kg and 160 g CP/kg, with a minimum of 0.80 g digestible Lys/MJ NE. Complete pedigree information was registered, starting at least one generation before F0 ancestors, to G10.

Selection candidates had records for feed intake and feed efficiency traits, growth traits, and live body composition traits. Response animals had records for the same traits recorded from 10 weeks of age until slaughter weight, plus carcass composition traits. In all generations boars were selected based on a phenotypic index combining daily feed intake (DFI) and average daily gain (ADG) between 35 and 95 kg BW, and backfat thickness (BFT) at 95 kg BW (Gilbert et al., 2007), as DFI (g/d) − (1.06 × ADG (g/d)) − (37 × BFT (mm)).

For the candidates to selection and the response animals, an RFI was computed as the residual of a multiple linear regression applied to DFI, using realized phenotypic correlations with traits accounting for production requirements (growth rate and body composition) and maintenance requirements (average metabolic BW (AMBW)), and the fixed effects of sex, pen size, CG, and the covariate of BW at the beginning of the test for response animals (Gilbert et al., 2017). Different equations were used for the two groups of animals, to account for the test differences. The RFI equation for selection candidates included ADG and BFT (measured by ultrasound), but because the test was run between fixed BW, AMBW would be equal for all animals and therefore was skipped from the equation. For response animals, the RFI equation included AMBW, ADG, carcass BFT (carcBFT) and lean meat content (LMC; computed from cut weights). Feed conversion ratio (FCR) was computed based on the corresponding test period of the two groups of animals.

In this study, five phenotypic traits available in both types of animals were studied: RFI, FCR, DFI, ADG, and BFT. To increase the statistical power, given the high r_g estimated in preliminary analyses between the traits measured in candidate and response animals, the phenotypic records were combined for both cohorts, after standardization of the records from candidates to selection to the variance of the corresponding trait of the

response animals, as describe Aliakbari et al. (2020). Descriptive information of the five traits from G0 to G10 are given in Table 3-1.

Table 3-1. Number (N), minimum (Min), maximum (Max), mean and standard deviation (SD) of the studied traits† in the low residual feed intake (LRFI) and high RFI (HRFI) lines

		N	Min	Max	Mean	SD	P-value‡
RFI	LRFI	1901	-0.38	0.37	-0.04	0.12	***
	HRFI	1748	-0.33	0.39	0.05	0.11	
FCR	LRFI	2190	1.60	3.88	2.61	0.25	***
	HRFI	1981	2.13	3.93	2.82	0.27	
DFI	LRFI	2172	1.25	2.92	2.02	0.25	***
	HRFI	1974	1.37	2.97	2.19	0.27	
BFT	LRFI	2058	9.82	44.63	25.44	7.01	***
	HRFI	1863	9.67	46.76	26.45	7.44	
ADG	LRFI	2251	0.51	1.02	0.76	0.08	*
	HRFI	2060	0.50	1.01	0.76	0.08	

†*ADG* average daily gain (kg/day), *BFT* backfat thickness (mm), *DFI* daily feed intake (kg/day), *FCR* feed conversion ratio (kg/kg), *RFI* residual feed intake (kg/day)

‡P-value of the effect of line in a linear model

3.2.3.2. Faeces sampling, microbial DNA extraction, 16S rRNA gene sequencing and sequence pre-processing

The microbiota information is most often derived from partial sequencing of the bacterial 16S ribosomal RNA (rRNA) gene, a housekeeping gene in all bacteria (Woese, 1987). Sequencing the 16S rRNA gene has become a standard approach in bacterial taxonomic classification, due to its ease to generate phylogenetic information at high throughput (Wang et al., 2015). For this purpose, nine hypervariable regions (V1-V9) of the 16S rRNA

gene can be targeted for sequencing. Sequences can then be analysed as separate Amplicon Sequence Variant (ASV), or clustered into 'Operational Taxonomic Units' (OTUs) based on their similarities. The ASV approach enables easier comparison between studies (Callahan et al., 2017). These units allow inferring the taxonomy of species present in the targeted biological samples using several reference databases. The counts of each OTU or ASV throughout the samples form a matrix called abundance table that is the basis of downstream analyses. Faecal sampling is a convenient and non-invasive sampling method that provides a reasonably good representation of the gut microbial communities (Ingala et al., 2018). It is now more common than other sampling locations for profiling of microbial communities in large mammalian animal populations.

For our study, faecal samples of 604 animals from G9 and G10 of the LRFI and HRFI lines were collected at 15 weeks of age, homogenized and placed immediately in dry ice, before storage at -80° C. The animals collected in G9 were the boars candidate to selection, and the pigs in G10 were females and castrated males response to selection. Microbial profiling was done as described previously (Achard et al., 2020). Briefly, the microbial DNA was extracted using the Quick-DNA™ Faecal Microbe Miniprep Kit™ (Zymo Research, Freiburg, Germany) and a 15 min bead-beating step at 30 Hz was applied. The V3-V4 region was then amplified from diluted genomic DNA with the primers F343 (CTTTCCCTACACGACGCTCTTCCGATCTTACGGRAGGCAGCAG) and R784 (GGAGTTCAGACGTGTGCTCTTCCGATCTTACCAGGGTATCTAATCCT) using 30 amplification cycles with an annealing temperature of 65 °C. This V3-V4 region has proved useful to study the variability of the pig microbiota in previous studies (Le Floc'h et al., 2014; L. M. G. Verschuren et al., 2018). The ends of each read overlap and can be stitched. In a single run, it generates extremely high quality, full-length reads of the full V3 and V4 region. The Flash software v1.2.6 (Magoc & Salzberg, 2011) was used to assemble each pair-end sequence, with at least a 10-bp overlap between the forward and reverse sequences, allowing 10% mismatch. Single multiplexing was performed using an in-house 6 bp index, which was added to R784 during a second PCR with 12 cycles using forward primer (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC) and reverse primer (CAAGCAGAAGACGGCATACGAGAT-index-GTGACTGGAGTTCAGACGTGT). The resulting PCR products were purified and loaded

to the Illumina MiSeq cartridge following the manufacturer's instructions. Run quality were internally checked using PhiX, and each pair-end sequence was assigned to its sample using the integrated index, with the bcl2fastq Illumina software. The sequences were submitted to the Short-Read Archive with accession number SRP124929. Filtering and trimming of sequences of high quality was applied to the reads with the DADA2 package in the R software (Callahan et al., 2016) with the following parameters: maxN=0, maxEE=2, truncQ=2, trimleft=17. Chimera were removed with the consensus method to obtain the final OTU abundance table. No further clustering was applied, so OTU were equivalent to ASV in this study. This step was followed by taxonomic annotation using the assignTaxonomy function of dada2 with the Silva Dataset v132 (Quast et al., 2013).

The final abundance table was rarefied to 9000 counts per sample, and contained 6792 OTUs or 298 genera across 604 samples. The 16 samples that contained fewer reads than 9000 were discarded, resulting in 588 samples in the final abundance table, 295 LRFI and 293 HRFI pigs. The microbiota analyses were then run at the genus level. The OTU relative abundances with the same taxonomic path until an identical genus were thus aggregated in a single count. Counts belonging to unclassified genera of a family were systematically gathered into a pseudo genus named NA_Family.

Finally, to limit the deviations of the genera distribution from the Gaussian distribution assumption used in linear mixed models (see next section), the genera table was filtered for a maximum proportion of 20% zero abundancy for each genus, and the resulting abundancies were log-transformed after adding a constant value of 1 to all counts. After this filtration step, 75 genera remained for the downstream analyses.

3.2.3.3. Statistical analyses

The beta-diversity is usually used to demonstrate the community differentiation between cohorts (Whittaker, 1960). To represent the beta-diversity between the faecal microbial genera communities of both lines, a non-metric multidimensional scaling (NMDS) based on the Bray-Curtis dissimilarity distance matrix was applied to the abundance table. This analysis was done using the R software and package "vegan" (Oksanen et al., 2013). The individual loadings were retrieved for each sample for the two first dimensions of the

NMDS. Then, the line effect was tested with a generalized linear model (GLM) on the loadings of the first two dimensions of the NMDS, the α-diversity metrics, the genera abundances, and the production traits. In addition, contributions of the genera to each axis, and to the plan defined by the two first axes, were computed as the squares of the loadings and sum of squares of the loadings, respectively. Before tests for line differences, variables with positive values (counts and diversity indexes) were log-transformed, whereas the loadings of the NMDS that contained negative values were submitted to a Johnson transformation (Johnson, 1949). These analyses were performed using package “car” in the R software (Fox et al., 2012) and the line effect was declared significant for $p < 0.05$ for the corresponding F-test.

To better understand how the genera are distributed, two α-diversity metrics, the Shannon (Shannon, 1948) and Simpson (Simpson, 1949) metrics, were calculated from the filtered table with 75 genera, and analysed as additional individual microbial traits. Following the main objective of the study, searching for the genetic relationships between microbiota traits and FE traits required the estimations of (co)variance components. The best linear unbiased prediction (BLUP) method was applied to the filtered genera and the two α-diversity metrics to obtain the (co)variance components. To follow the assumption of the BLUP method, which should be applied to a non-selected base population, all analyses were done in bivariate models including the selection index as the first trait. The second trait was the microbiota observation vector (abundance of each genus or α-diversity metric). To compute genetic correlations between the performance traits and microbiota observations, each of the production traits was added in three-variate analyses.

The significance of fixed environmental factors ($p < 0.05$) on all response variables was tested in preliminary GLM analyses. Significant fixed factors, including pen size (5 levels), herd of birth (two levels), sex (three levels), and contemporary groups (CG, 109 levels) for performance traits, microbiota data and α-diversity metrics, were systematically fitted. The fitted covariates were slaughter body weight (BW) for BFT and BW at test for genera abundancies and α-diversity metrics. The significance of all fitted fixed factors on the 75 genera are given in Table S1. The litter effect was fitted as a random environmental

source of variation for performance traits, and for microbiota data whenever it was significant ($p < 0.05$ for a Chi^2 test applied to the likelihood ratio test comparing the models with and without this term).

The following bivariate and three-variate animal models were used to estimate the variance components:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_1a} + \mathbf{Z_2l} + \mathbf{e}$$

where **y** is the vector of observations for the index and the abundance of each genus or an α-diversity metric, and one of five performance traits (in three-variate analyses), **b** is the vector of fixed effects (described above), **a** is the vector of additive genetic effects, **l** is the vector of litter effects, and **e** is the vector of random residuals. **X**, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices for **b**, **a** and **l**. The distributions assumed for the random terms were $\mathbf{a} \sim N(0, \mathbf{G_0} \otimes \mathbf{A})$, $\mathbf{l} \sim N(0, \mathbf{R_l} \otimes \mathbf{I})$, and $\mathbf{e} \sim N(0, \mathbf{R_e} \otimes \mathbf{I})$, where $\mathbf{G_0}$ is a 2 × 2 or 3 × 3 symmetrical direct additive genetic effect (co)variance matrix, and $\mathbf{R_l}$ and $\mathbf{R_e}$ are 2 × 2 or 3 × 3 symmetrical litter effect and residual effect (co)variance matrices, respectively. **I** denoted the identity matrix of adequate dimension. The pedigree relationship matrix (**A**) included 10 generations of pedigree information plus ancestors, and contained 7293 animals. The analyses were performed using AIREMLF90 software (Misztal et al., 2018) for the BLUP method.

To test the significance of h^2 of the 75 genera, an empirical significance threshold equal to 0.125 was considered. The threshold was obtained after running 10000 univariate analyses using the above described genetic model applied to microbiota abundancies, based on a null hypothesis of no genetic control on the abundancies. The null hypothesis was obtained by shuffling the abundances across individuals for two arbitrary genera. The minimum value of the top 5% of the estimated h^2 was considered as the threshold to decide that a genus was heritable. Thereafter, the three-variate analyses were conducted for genera with h^2 significantly different from zero. The deviation from zero of the additive r_g of genera and α-diversity metrics with the production traits were tested using a Z-test.

3.2.4. Results

3.2.4.1. Gut microbiome differences between lines

The 75 filtered genera represented on average 97% of the sample counts of the rarefied table. Among these genera, 42 had significantly higher abundances in the LRFI line than in the HRFI line, and 10 were more abundant in the HRFI line (Figure 3-1 and Table S2).

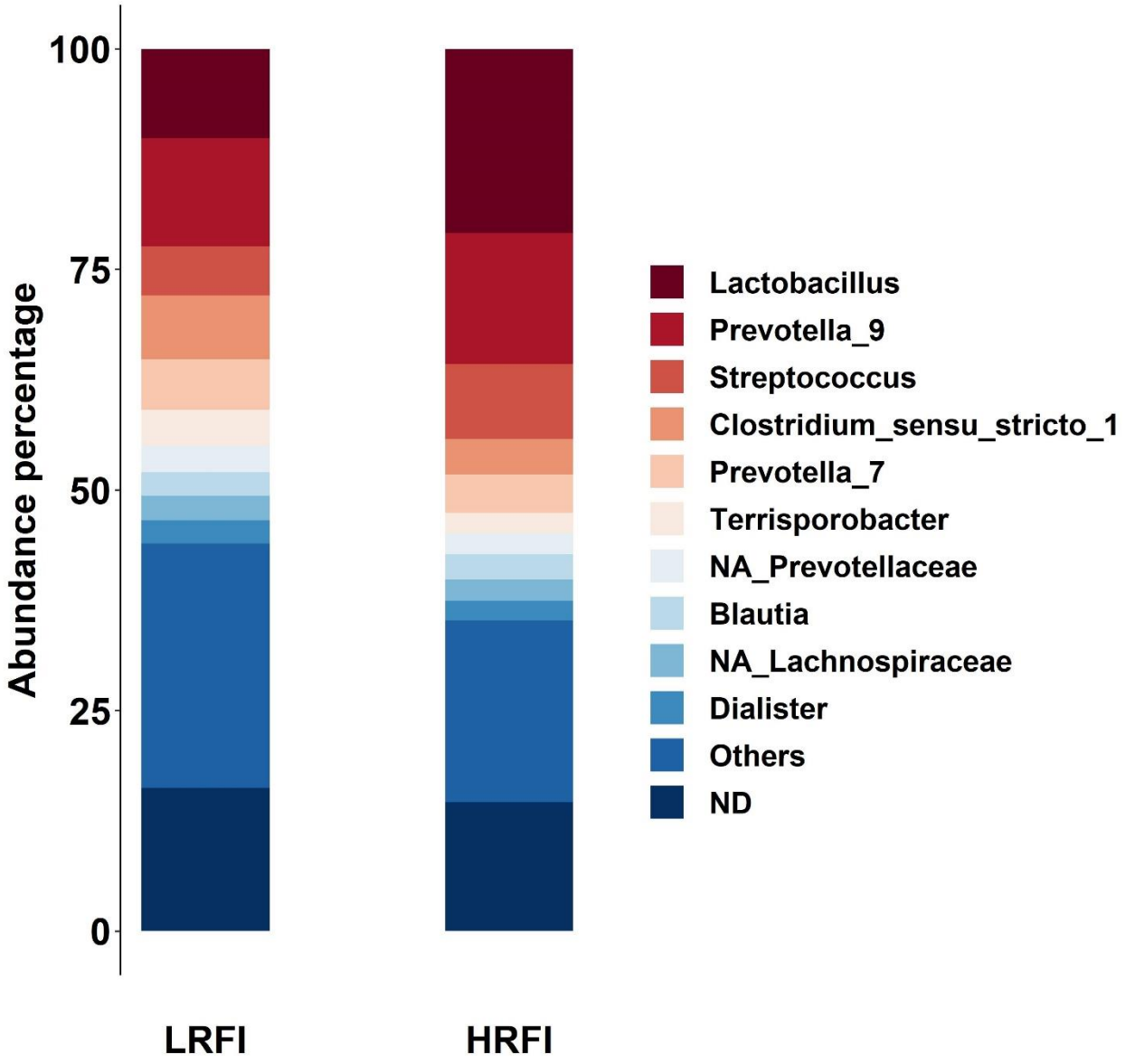


Figure 3-1. Abundance percentage of the 75 genera in the LRFI and HRFI lines.

Others **=** differentially abundant genera between lines with abundances lower than 2%, ***ND*** = genera with non-significant abundance difference between the two lines

Of the differentially abundant genera between lines (*P*-value < 0.05 for a Student test applied to the log-transformed abundances), the genera *Lactobacillus* (~10.1% in the LRFI line and ~20.9% in the HRFI line of the 75 genera counts, *P*-value < 0.0001), *Prevotella_9* (~12.2% and ~14.8% in the LRFI and HRFI lines, respectively, *P*-value < 0.03), and *Streptococcus* (~5.6% in the LRFI line vs ~8.5% in the HRFI line, *P*-value < 0.0001) were the more abundant genera in both lines, and they were all more abundant in the HRFI line. The three genera *Clostridium_sensu_stricto_1* (*P*-value < 0.0001), *Prevotella_7* (*P*-value < 0.004), and *Terrisporobacter* (*P*-value < 0.0001) were more abundant in the LRFI line (~7.2%, ~5.7% and ~4.1%, respectively) than in the HRFI line (~4.0%, ~4.4% and ~2.3%, respectively). The four genera *Dialister* (*P*-value < 0.05), *NA_Prevotellaceae* (*P*-value < 0.0001), *NA_Lachnospiraceae* (more abundant in the LRFI line, *P*-value < 0.0001), and *Blautia* (more abundant in the HRFI line, *P*-value < 0.0001) represented on average ~2.2% of the counts. The other 42 differentially abundant genera had abundances lower than 2% in the two lines, and represented a total of 25.9% and 18.6% of the abundances in the LRFI and HRFI lines, respectively. The remaining 23 genera that were not significantly different (*P*-value > 0.05) between the lines had total abundance of ~16.2% in the LRFI and ~14.6% in the HRFI lines.

The NMDS showed differences between the genera communities of the LRFI and HRFI lines (Figure 3-2). The two lines were significantly differentially distributed only along the second ($p < 0.01$) dimension. Among the 75 genera included in the NMDS, the genus *Rikenellaceae_RC9_gut_group* had the highest (2.9%) contribution to the plan defined by dimensions 1 and 2, and the genus *Succinivibrio* had the lowest (0.03%) contribution (Table S2 and Figure 3-3). In details, on the first axis 25 genera had a contribution larger than the expected contribution if all genera contributed equally 1.33% (100/75), including 15 differentially abundant between the lines. It was mainly driven (contributions larger than 3.2%) by the opposition of the genera *Prevotella_7* (5.2%), *Syntrophococcus* (5.0%), *NA_Family_XIII* (5.0%), *Lachnospiraceae_NK3A20_group* (4.7%), *Olsenella* (4.6%), *Dialister* (4.5%), *Mitsuokella* (4.5%), and *Shuttleworthia* (4.3%) in one direction, and the genera *Lachnospiraceae_ND3007_group* (4.0%), *Ruminococcaceae_UCG-008* (3.9%), and *Marvinbryantia* (3.4%) in the other direction. On the second axis, 25 genera had contributions larger than 1.33%, including 22 genera

differentially abundant between the lines. The genera *Prevotella_9* (4.3%) drove the direction towards more HRFI samples, whereas the genera *Ruminococcaceae_NK4A214_group* (5.7%), *Rikenellaceae_RC9_gut_group* (5.6%), *Ruminococcaceae_UCG-002* (5.1%), *Family_XIII_AD3011_group* (4.5%), *NA_Ruminococcaceae* (4.4%), *Christensenellaceae_R-7_group* (4.2%), *NA_Muribaculaceae* (4.0%), *Ruminococcaceae_UCG-005* (3.6%), *Prevotellaceae_UCG-001* (3.3%) and *Ruminococcaceae_UCG-010* (3.2%) were the main contributors to the opposite direction, toward the LRFI line (contributions higher than 3.2%).

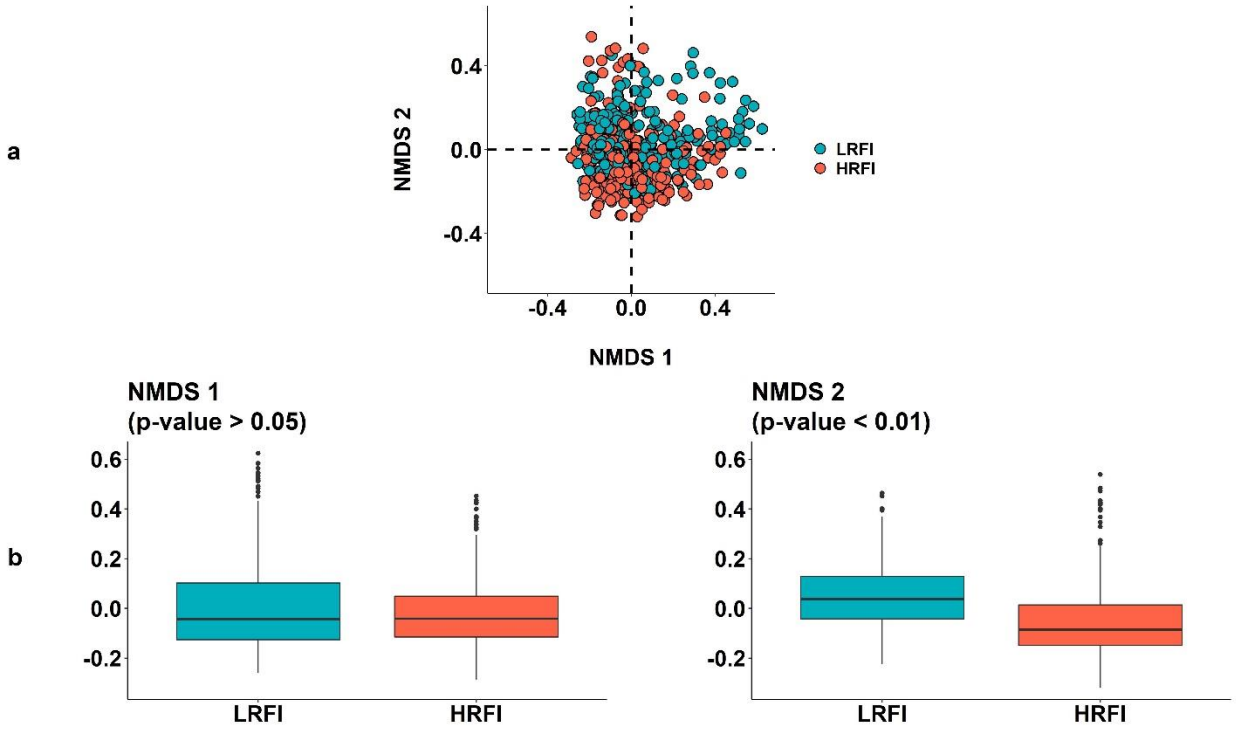


Figure 3-2. Non-metric multidimensional scaling (NMDS) based on the Bray-Curtis dissimilarity matrix of the genera community (a) and box plots of the individual coordinates per line on the two first axes of the NMDS (LRFI=low residual feed intake; HRFI= high residual feed intake), with P-value of the ANOVA test of the line differences (b)

The Shannon and Simpson α-diversities indexes showed significantly higher microbial diversity in the LRFI line than in the HRFI line ($p < 0.01$, Figure 3-4).

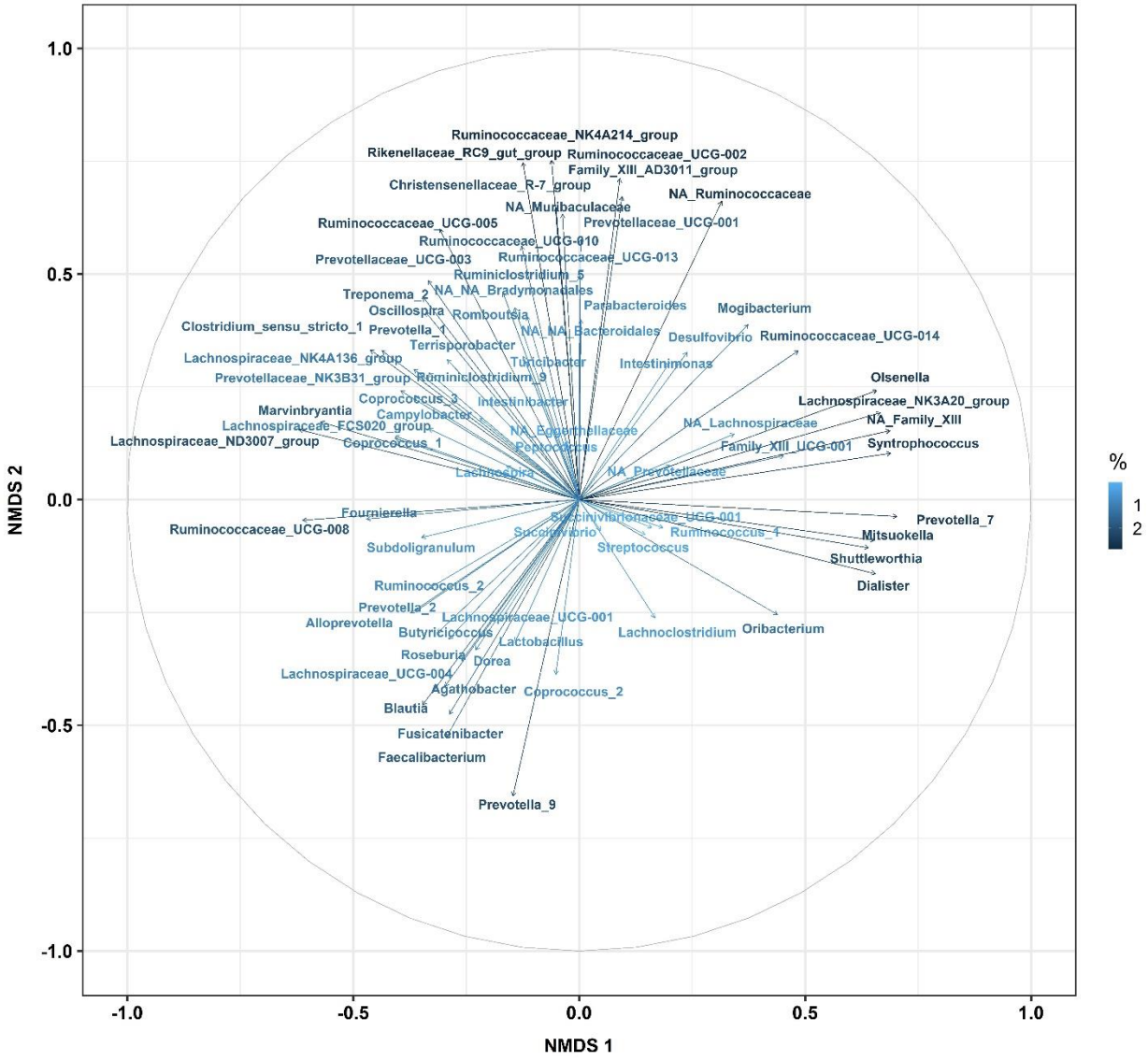


Figure 3-3. Projection of the genera on the first and second dimensions in a non-metric multidimensional scaling (NMDS) applied to the Bray-Curtis matrix of the genera abundances. The arrows are colored based on the contribution of each genus to the plan.

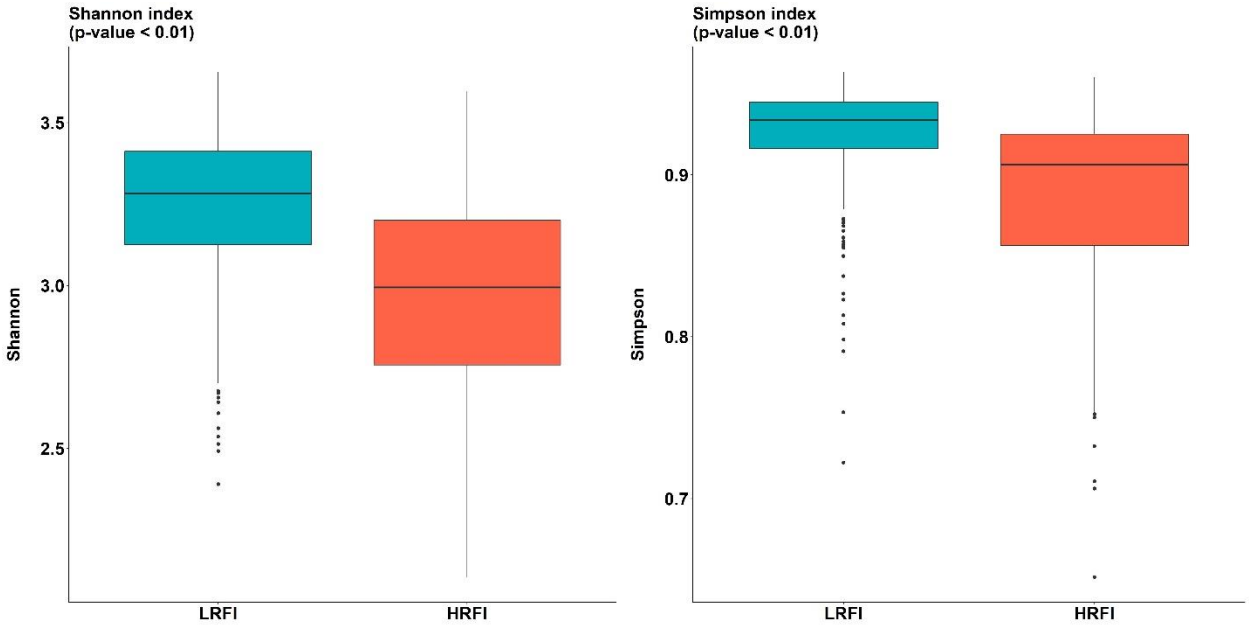


Figure 3-4. Box plots of Shannon and Simpson α-diversity indexes per line

(LRFI=low residual feed intake (n=295); HRFI= high residual feed intake (n=293)) and P-value of ANOVA test of the line differences

3.2.4.2. Heritability estimates of microbiota traits

The gut microbiota composition can be highly heritable in pigs, but not for all genera. The h^2 estimates for the Shannon and Simpson α-diversities indexes were 0.19 ± 0.08 and 0.12 ± 0.06, respectively (Table 3-2). The estimated h^2 of the genera ranged from null to 0.50 ± 0.12 for *Clostridium_sensu_stricto_1.* Forty-eight genera had a h^2 higher than 0.125, and therefore were considered as heritable, including 34 genera with h^2 larger than 0.20. The majority of the genera that were differentially abundant between lines were heritable (33/52). Out of the 23 genera that did not differ between lines, 15 had significant h^2. For the 48 heritable genera, the abundances per line are shown in Table S2 and Figure 3-5.

Heritable genera were also more abundant genera, while non-heritable genera tended to be at lower abundance (*P*-value < 0.05 for a Student test applied to the average of log-transformed abundances). A Spearman correlation of 0.26 (*P*-value < 0.05) was

estimated between the h^2 estimates and the average of log-transformed abundances, while a correlation of 0.10 (*P*-value > 0.05) was obtained with the raw averages.

Comparison of the contributions of the heritable and non-heritable genera to the axes of NMDS showed a significant difference (*P*-value < 0.05) of contribution to the first axis between the two groups of genera: the average contribution of the heritable genera to axis 1 was 1.8%, whereas the non-heritable genera had an average contribution of 0.5%. The two groups of genera similarly contributed to the second axis (*P*-value = 0.08): the average contribution of the heritable and non-heritable genera to the second axis were 1.1% and 1.8%, respectively.

3.2.4.3. Genetic correlations of microbiota traits with production traits

The two α-diversities indexes and 48 genera with significant h^2 were included in three-variate analyses to estimate genetic correlations with production traits. The r_g of the α-diversities indexes with production traits were negative and similar for the two metrics (Table 3-2). With ADG, DFI, and RFI r_g estimates were lower than 0.27, and did not differ from zero. The highest r_g were obtained with BFT (r_g < -0.89 ± 0.04) and FCR (-0.61 ± 0.52).

Table 3-2. Estimated heritability (h^2) and standard errors (SE) and descriptive statistics (minimum (Min), maximum (Max), mean and standard deviation (SD)) of α-diversity indexes and genera abundances

	h^2 ± SE†	% Zeros	Min	Max	Mean	SD
α-diversity index						
Shannon	0.19 ± 0.08	0	2.1	3.6	3.1	0.3
Simpson	0.12 ± 0.06	0	0.6	1.0	0.9	0.05
Genus						
Clostridium_sensu_stricto_1	0.50 ± 0.12	0	2	2574	488.6	442.9
Prevotella_1	0.44 ± 0.11	14	0	306	35.3	45.5
Blautia	0.39 ± 0.11	0	14	571	241.5	99.9
Prevotellaceae_NK3B31_group	0.36 ± 0.10	2	0	1216	84.6	131.1
Lachnospiraceae_NK3A20_group	0.36 ± 0.10	2	0	1663	77.7	186.8
Ruminococcaceae_UCG-008	0.35 ± 0.11	2	0	307	89.1	58.7

Lachnospiraceae_ND3007_group	0.35 ± 0.11	3	0	109	24.4	17.0
Coprococcus_3	0.35 ± 0.10	1	0	432	48.4	37.1
Butyricicoccus	0.34 ± 0.11	4	0	68	18.2	12.4
Terrisporobacter	0.34 ± 0.11	0	13	1070	279	193.7
Syntrophococcus	0.34 ± 0.10	6	0	1480	74.4	119.9
Faecalibacterium	0.33 ± 0.11	0	1	527	169.1	98.3
Coprococcus_1	0.32 ± 0.10	13	0	257	14.7	24.4
Marvinbryantia	0.30 ± 0.10	5	0	141	20.7	17.9
Mitsuokella	0.30 ± 0.09	1	0	1083	93.6	110.7
NA_Family_XIII	0.29 ± 0.09	5	0	285	24.1	36.2
Prevotella_7	0.28 ± 0.10	0	12	2583	441.7	335.5
Prevotellaceae_UCG-003	0.28 ± 0.10	3	0	240	18.9	24.4
Romboutsia	0.28 ± 0.10	4	0	235	36	39.9
Fusicatenibacter	0.27 ± 0.10	3	0	117	24.8	16.8
Campylobacter	0.27 ± 0.10	4	0	266	23.6	26.9
Olsenella	0.27 ± 0.09	8	0	735	42.6	87.5
Oscillospira	0.25 ± 0.09	14	0	85	9.7	10.6
Lactobacillus	0.24 ± 0.09	0	17	5034	1353.5	1148.3
Roseburia	0.23 ± 0.10	2	0	346	79.5	64.2
Succinivibrionaceae_UCG-001	0.23 ± 0.09	13	0	1153	94.5	161.2
NA_Muribaculaceae	0.23 ± 0.08	0	0	727	68	73.1
Dorea	0.22 ± 0.09	1	0	648	67.8	47.8
Subdoligranulum	0.22 ± 0.09	0	11	583	175.2	93.2
Alloprevotella	0.22 ± 0.09	0	3	389	86.1	54.8
Ruminococcaceae_UCG-014	0.22 ± 0.09	0	3	505	108.9	71.9
Dialister	0.20 ± 0.08	0	3	844	213	135.5
Shuttleworthia	0.20 ± 0.09	0	1	1543	280.7	250.3
Streptococcus	0.20 ± 0.10	0	20	2526	613.1	446.4
NA_Prevotellaceae	0.20 ± 0.09	0	2	817	233.6	102.9
Rikenellaceae_RC9_gut_group	0.20 ± 0.09	0	0	862	118.3	116.0
Lachnospiraceae_NK4A136_group	0.19 ± 0.08	5	0	253	21.6	23.0
Desulfovibrio	0.18 ± 0.09	1	0	192	21.5	20.8
Lachnospiraceae_UCG-001	0.18 ± 0.08	14	0	83	12	12.7
Ruminococcus_2	0.17 ± 0.09	6	0	168	31.1	26.8
NA_Ruminococcaceae	0.16 ± 0.08	0	19	749	135.5	89.3
Treponema_2	0.16 ± 0.08	7	0	761	58.4	98.2
Fournierella	0.14 ± 0.08	13	0	66	9	9.1

Prevotella_2	0.14 ± 0.08	0	0	340	75	56.0
Agathobacter	0.14 ± 0.07	0	5	823	253.7	148.1
Lachnospira	0.13 ± 0.07	1	0	263	44	33.0
Ruminococcaceae_UCG-005	0.13 ± 0.07	0	0	665	72.6	73.2
Lachnospiraceae_UCG-004	0.13 ± 0.07	16	0	25	5.3	4.6
Ruminococcaceae_UCG-013	0.12 ± 0.07	14	0	73	7.8	8.6
Intestinimonas	0.10 ± 0.06	5	0	38	7.5	5.5
Turicibacter	0.10 ± 0.03	6	0	246	31.7	37.2
Intestinibacter	0.09 ± 0.08	0	1	258	39.5	24.3
Oribacterium	0.09 ± 0.06	1	0	151	42.9	24.5
Ruminiclostridium_5	0.08 ± 0.07	6	0	47	8.1	6.4
Family_XIII_AD3011_group	0.08 ± 0.06	1	0	303	37	33.7
Christensenellaceae_R-7_group	0.07 ± 0.06	1	0	933	52.3	99.8
Lachnospiraceae_FCS020_group	0.07 ± 0.06	2	0	43	11.1	6.8
NA_NA_Bradymonadales	0.06 ± 0.06	19	0	356	23.1	37.6
Family_XIII_UCG-001	0.06 ± 0.06	2	0	45	15.2	8.9
Mogibacterium	0.06 ± NE‡	5	0	130	12.9	12.6
Succinivibrio	0.05 ± 0.05	5	0	501	29.6	46.1
NA_Eggerthellaceae	0.05 ± 0.06	10	0	30	6.2	5.0
Ruminiclostridium_9	0.04 ± 0.05	7	0	36	7.3	5.8
Lachnoclostridium	0.04 ± 0.05	7	0	140	11.8	11.7
Ruminococcaceae_NK4A214_group	0.03 ± 0.01	1	0	244	36.4	35.2
Ruminococcaceae_UCG-002	0.02 ± 0.01	0	1	584	69.4	67.5
NA_Lachnospiraceae	0.02 ± 0.01	0	62	661	226	76.3
Ruminococcaceae_UCG-010	0.02 ± 0.01	2	0	970	40.9	76.1
Prevotellaceae_UCG-001	0.01 ± NE	19	0	128	8	14.5
Prevotella_9	0.01 ± NE	0	34	2935	1180.6	561.1
NA_NA_Bacteroidales	0.01 ± NE	5	0	204	15.9	24.1
Coprococcus_2	0 ± NE	7	0	81	15.4	12.9
Peptococcus	0 ± NE	2	0	62	14	8.0
Ruminococcus_1	0 ± NE	0	27	408	143.1	51.9
Parabacteroides	0 ± NE	12	0	247	12.8	21.6

†h^2 were obtained after log transformation, ‡NE: not estimable

With the genera, r_g ranged from -0.36 ± 0.24 (*Romboutsia*) to 0.32 ± 0.12 (*Streptococcus*) with RFI, from -0.38 ± 0.55 (*Ruminococcaceae_UCG-005*) to 0.51 ± 0.31 (*Fusicatenibacter*) with FCR, from -0.63 ± 0.45 (*Desulfovibrio*) to 0.60 ± 0.12

(*Faecalibacterium*) with DFI, -0.98 ± NE (*NA_Ruminococcaceae*) to 0.86 ± 0.05 (*Lactobacillus*) with BFT, and from -0.48 ± 0.56 (*NA_Ruminococcaceae*) to 0.73 ± 0.76 (*Lachnospiraceae_UCG-001*) with ADG. In Table 3-3, the r_g of the 22 genera that had at least one significant genetic correlation with the performance traits are presented. The production trait with the highest number of significant r_g with genera was BFT (11 significant correlations with genera). In addition, three genera had r_g estimates close to -1 with this trait (*Desulfovibrio*, *NA_Ruminococcacaea*, *Lachnospira*), but Z-tests could not be applied for these cases, as standard errors were not estimable at the borders of the parameter space. The DFI and RFI showed significant r_g with 7 and 3 genera, respectively, and there were no genera with significant r_g with ADG and FCR. The genus *Shuttleworthia* had significant genetic correlations with two traits (DFI and BFT), and the genus *Desulfovibrio* had a significant r_g with RFI and close to -1 with BFT.

From the 10 genera more abundant in the HRFI line, 6 had significant r_g with at least one production trait, and out of the 42 genera more abundant in the LRFI line, only 7 had significant correlations with the production traits. The other 9 genera with significant genetic correlations with at least one trait were from the 23 genera that had similar abundances between the lines. Distribution between the LRFI and HRFI lines of the abundance of the 22 genera with significant r_g are presented in Figure 3-5. The three genera with significant r_g with RFI (*Streptococcus*, *Desulfovibrio*, and *Prevotella_2*) had significant line abundance differences that were consistent with the sign of the r_g. The genera *Streptococcus* and *Prevotella_2* were more abundant in the HRFI line and had a positive r_g with RFI, whereas the genus *Desulfovibrio* was more abundant in the LRFI line, and had a negative r_g with RFI. Out of the 14 genera with significant or very negative genetic correlations with BFT, genera *Blautia*, *Lactobacillus*, and *Dorea* were significantly more abundant in the HRFI line, and had positive r_g with BFT, and the 5 genera *Prevotella_7*, *Rikenellaceae_RC9_gut_group*, *Desulfovibrio*, *NA_Ruminococcaceae,* and *Lachnospira* were more abundant in the LRFI line, and had negative r_g with BFT. Of the 7 genera that had significant r_g with DFI, only the genus *Roseburia* (more abundant in the LRFI line) had significant abundance difference between the two lines, and the sign of the r_g was not consistent with the line differences.

Box plots showing genera abundances between the LRFI and HRFI lines for the other 53 genera are given in supplementary Figure S1.

Table 3-3. Genetic correlations[†] (SE) of α-diversity indexes and genera with production traits[‡]

	RFI	FCR	DFI	BFT	ADG
α-diversity index					
Shannon	-0.26 ± 0.29	-0.61 ± 0.52	-0.30 ± 0.29	-0.89 ± 0.04*	-0.21 ± 0.32
Simpson	-0.27 ± 0.34	-0.93 ± NE[§]	-0.42 ± 0.34	-0.94 ± NE	-0.31 ± 0.48
Genus					
Blautia	0.20 ± 0.12	0.32 ± 0.23	0.33 ± 0.25	0.50 ± 0.22*	0.02 ± 0.26
Ruminococcaceae_UCG-008	0.05 ± 0.23	0.26 ± 0.23	0.32 ± 0.23	0.54 ± 0.22*	-0.01 ± 0.28
Coprococcus_3	-0.03 ± 0.24	0.27 ± 0.22	0.25 ± 0.21	0.56 ± 0.21*	-0.12 ± 0.27
Syntrophococcus	-0.04 ± 0.25	-0.18 ± 0.26	-0.29 ± 0.23	-0.60 ± 0.23*	-0.03 ± 0.28
Faecalibacterium	0.20 ± 0.12	0.26 ± 0.30	0.60 ± 0.12*	0.41 ± 0.33	0.18 ± 0.32
Coprococcus_1	-0.09 ± 0.25	0.18 ± 0.25	0.30 ± 0.23	0.54 ± 0.24*	0.12 ± 0.29
Marvinbryantia	0.10 ± 0.24	0.19 ± 0.25	0.29 ± 0.28	0.47 ± 0.24*	-0.04 ± 0.29
Prevotella_7	-0.19 ± 0.13	-0.11 ± 0.27	-0.28 ± 0.32	-0.71 ± 0.28*	-0.08 ± 0.31
Lactobacillus	0.29 ± 0.24	-0.05 ± 0.19	0.51 ± 0.34	0.86 ± 0.05*	0.30 ± 0.35
Roseburia	0.01 ± 0.14	-0.05 ± 0.32	0.35 ± 0.12*	0.16 ± 0.50	0.31 ± 0.65
Dorea	0.14 ± 0.16	0.05 ± 0.47	0.33 ± 0.43	0.66 ± 0.29*	0.10 ± 0.40
Shuttleworthia	-0.13 ± 0.14	-0.05 ± 0.34	-0.51 ± 0.10*	-0.76 ± 0.36*	-0.28 ± 0.40
Streptococcus	0.32 ± 0.13*	-0.24 ± 0.31	-0.17 ± 0.13	-0.49 ± 0.39	-0.38 ± 0.57
Rikenellaceae_RC9_gut_group	-0.14 ± 0.29	-0.12 ± 0.37	-0.43 ± 0.38	-0.86 ± 0.06*	-0.45 ± 0.44
Desulfovibrio	-0.30 ± 0.13*	-0.35 ± 0.65	-0.63 ± 0.45	-0.97 ± 0.01[NE]	-0.30 ± 0.52
Lachnospiraceae_UCG-001	-0.01 ± 0.33	-0.03 ± 0.36	0.55 ± 0.12*	0.39 ± 0.42	0.73 ± 0.76
Ruminococcus_2	0.08 ± 0.14	-0.14 ± 0.48	0.44 ± 0.12*	0.14 ± 0.49	0.18 ± 0.58
NA_Ruminococcaceae	-0.16 ± 0.66	-0.18 ± 0.40	-0.54 ± 0.49	-0.98 ± 0.01 [NE]	-0.48 ± 0.56
Prevotella_2	0.30 ± 0.13*	0.49 ± 0.52	0.33 ± 0.64	0.59 ± 0.57	-0.09 ± 0.49
Agathobacter	0.24 ± 0.13	0.19 ± 0.37	0.59 ± 0.12*	0.16 ± 0.66	0.53 ± 0.65
Lachnospira	-0.03 ± 0.36	-0.15 ± 0.49	0.04 ± 0.34	-0.95 ± 0.02 [NE]	0.38 ± 0.45
Lachnospiraceae_UCG-004	0.09 ± 0.15	0.42 ± 0.44	0.47 ± 0.12*	0.14 ± 0.49	0.22 ± 0.66

[†]* indicate genetic correlations different from zero with a Z test ($P < 0.05$)

[‡†]*ADG* average daily gain, *BFT* backfat thickness, *DFI* daily feed intake, *FCR* feed conversion ratio, *RFI* residual feed intake, [§]NE: not estimable

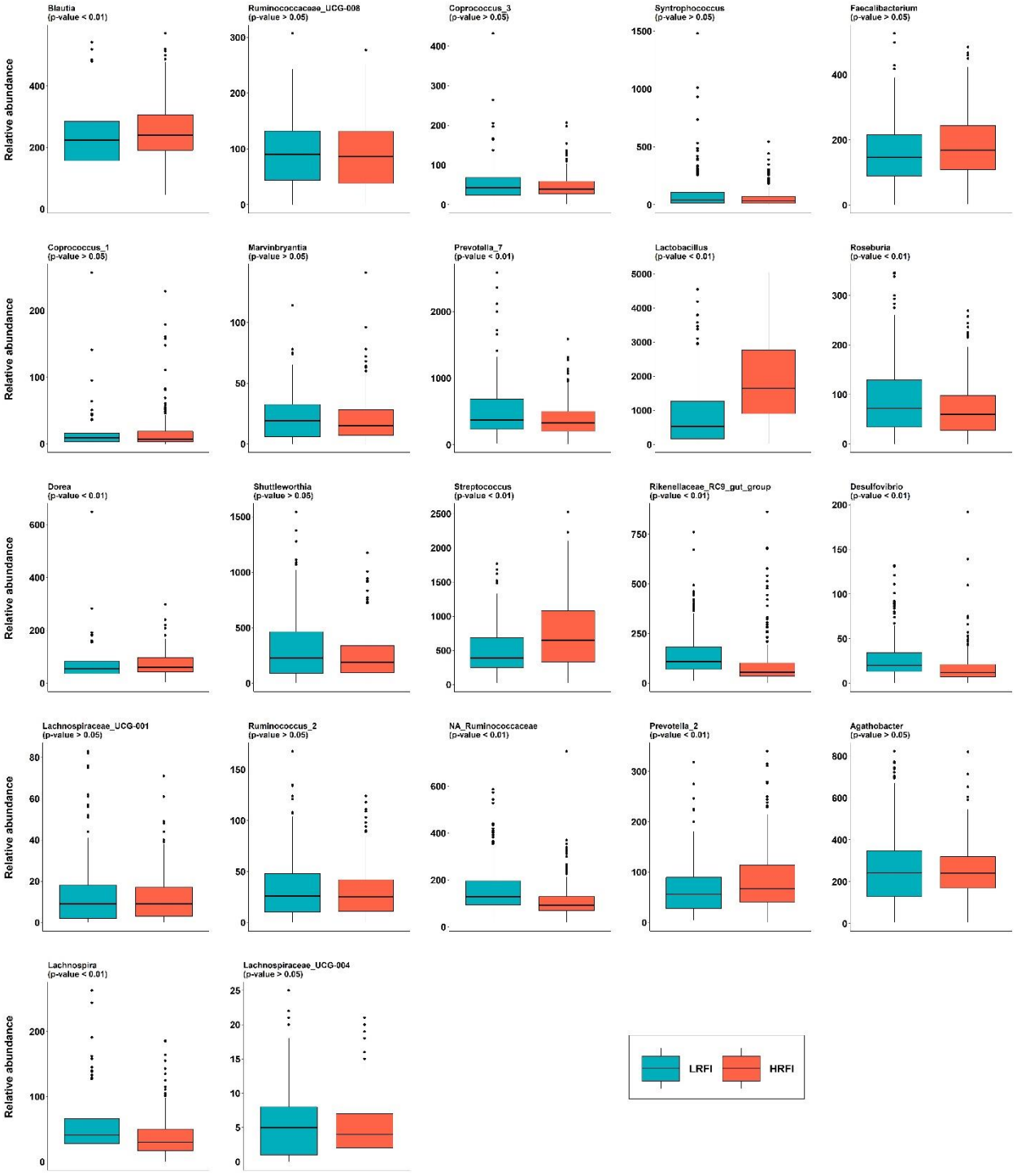


Figure 3-5. Box plots of genera abundances per line (LRFI, low residual feed intake; HRFI, high residual feed intake) and *p*-value of ANOVA test of the line differences

3.2.5. Discussion

The objective of the present study was to clarify if some components of pig faecal microbiota have genetic relationships with production and FE traits, taking advantage of data collected in two experimental pig lines divergent for RFI. The approach combined a comparison of the microbiota composition between the genetic lines, and quantitative genetic models to quantify the genetic control on the microbiota components and estimate genetic correlations with traits of interest. These approaches were applied to the subset of genera counts that presented reasonably good properties (number of zeros and Gaussian distribution) to be submitted to linear mixed models. A substantial genetic control for these genera abundances was evidenced with the two approaches, and interesting genetic relationships with the traits of interest were pointed out.

3.2.5.1. Some genera are under genetic control

Most studies that compared microbiome data of pigs between low and high RFI groups are based on a phenotypic selection of extreme pigs in a population, so most of the reported differences would be driven by phenotypic relationships. In our study differences between animals were established by at least 9 generations of selection, therefore a large proportion of the line differences would result from genetic differences between pigs. Because of the limited size of the lines, the differences can result from an association with the selected trait, or from genetic differences arising by chance (i.e. drift; Hill (1972)). The quantitative models that combine microbiota and production traits thus provide a complementary approach to evidence genetic relationships between FE and gut microbiota, but its power is more limited than line comparisons.

Some genera differentially abundant between lines pointed out to genera previously reported as associated with feed intake or feed efficiency. Among the most abundant genera that differed between lines, the genus *Lactobacillus* was one the more abundant ones, with higher abundance associated with high RFI. This genus is well described for its commonness and its important functions in gut health in animals (Dowarah et al., 2017; V. D. Valeriano et al., 2017). *Lactobacillus* is the most abundant member of the lactic acid producer bacteria, and is routinely used as a probiotic

supplement in the swine nutrition because of its enzymatic activities in the digestion and absorption process of the nutrients in the gut (Kim et al., 2007). Several species of this genus have been reported to have effects on the studied traits (Giang et al., 2011; Shon et al., 2005; Yu et al., 2008). *Lactobacillus* has been reported to be enriched in the faeces of more healthy pigs and positively correlated with feed efficient animals (Bergamaschi, Tiezzi, et al., 2020a; H. Yang et al., 2017). Considering the better health of the LRFI pigs (Chatelet et al., 2018), the lower abundance of *Lactobacillus* in this line was surprising. Conversely, in a study on the faecal microbiota at 80 days of age in Duroc pigs, the genus *Lactobacillus* was reported as one of the four dominant genera in pigs with high RFI from 90 to 160 days of age and not in their low RFI counterparts (Si et al., 2020), which is consistent with the lower abundance of this genus in the LRFI line in our study. Similarly, L. M. G. Verschuren et al. (2018) reported a lower abundance of some OTUs belonging to the *Lactobacillus* genus in low FE than high FE gilts, but the reverse for boars. Overall, the favourable functions of the *Lactobacillus* genus could be partially covered by other genera in the LRFI pigs that showed more diversity than the HRFI animals. *Prevotella*, including *Prevotella_9* and *Prevotella_7,* was the second genus differentially abundant between lines. Si et al. (2020) reported a slightly higher abundance for this genus in animals with low RFI (16.25%) in comparison to animals with high RFI (12.48%), which is in contrast with the higher abundance of the genus *Prevotella_9* in HRFI pigs in our study, but is consistent with the more abundant *Prevotella_7* found in the LRFI line. However, He et al. (2019) also reported a lower abundance of *Prevotella_9* in more feed efficient (15.07%) compared to less feed efficient (17.85%) pigs. The prevalence of members of the *Prevotella* genera is related to their enhancer role in the digestion ability and extracting nutrients from high fiber plants (Plummer et al., 2020). This complex and relatively diverse genus seems to contain multiple functions related to the sub-genera reported in the more recent studies that are not yet clearly identified. The genus *Streptococcus*, more prevalent in the HRFI line, is another member of the lactic acid producer bacteria (du Toit et al., 2014). U. M. McCormack et al. (2017) reported a 2-fold lower abundance of the genus *Clostridium_sensu_stricto_1* in low RFI pigs than high RFI pigs, which is in contrast with our observed higher abundance in the LRFI line.

The results of NMDS confirmed the hypothesis of changes in the intestinal microbial community as a result of selection for feed efficiency. Even though the genera contributions were consistent with their prevalence in the lines (for instance, the genera *Lactobacillus* and *Prevotella_9* had negative loadings on the second axis, which corresponded to the direction of the HRFI line), the extent of the contributions was not related to the abundance in the two lines. For instance, genera from the *Ruminococcaceae* family had an abundance lower than 2% in the LRFI line, but they were among the highest positive contributors to the second axis.

Our results showed significant additive genetic variance for 61% of the analysed genera. Overall, observing significant heritabilities for more than half of the analysed genera, which represented about 97% of the gut microbial communities, suggests that a considerable part of variability of the gut microbial community is under genetic control. However, some heritable genera were shown to differ between lines, but some differentially abundant genera were not heritable, and some heritable genera did not differ between the lines. This last situation could correspond to genera with limited genetic relationship with the selection criterion that would thus not respond to selection and be differentially abundant. The situation of genera that were differentially abundant between lines and not heritable in our study can be related to a limited power of our experimental design to estimate accurately the variance components: only h^2 estimates higher than 0.12 could be declared significant, so all genera with low heritability would be ignored in our results. Besides, the slight correlation between h^2 estimates and the average genera abundances found in our study is usually not expected and is assumed to be due to the dataset truncation (genera with more than 20% of zero were not analysed, which tend to be the lowest abundant) and consequently missing heritable genera with low abundances. Limited sequencing depth of the microbiota data would cause less precise quantification and the high proportion of zeros that result in imperfect analyses of genera with low abundancies.

Except in few cases, our h^2 estimates were in the range of previously published values for these genera (Camarinha-Silva et al., 2017; Chen et al., 2018). For instance, Chen et al. (2018) reported an h^2 of 0.26 for genus *Turicibacter*, that is higher than our

estimate (0.10 ± 0.03). Among the genera that now have sub-types (*Prevotella*, *Coprococcus* and *Ruminococcus*), we have obtained different h^2 values for the different types. For *Prevotella*, h^2 ranged from 0.44 ± 0.11 for *Prevotella_1* to 0.01 ± NE (*Prevotella_9*). Chen et al. (2018) have been reported an h^2 of 0.23 for the genus *Prevotella* and 0.22 for the genus *Coprococcus* that are in agreement with our estimations for the *Prevotella_7* (0.28 ± 0.10) and *Coprococcus_1* (0.32 ± 0.10). The estimated h^2 for genus *Lactobacillus* (0.24 ± 0.09) was higher than reported value (0.08) by Chen et al. (2018) and lower than the value (0.34) reported by Camarinha-Silva et al. (2017). We obtained same h^2 for the genus *Blautia* (0.39 ± 0.11) as Camarinha-Silva et al. (2017) (0.33 ± 0.14), and a slightly lower h^2 for the genus *Alloprevotella* (0.22 ± 0.09) than their report (0.34 ± 0.16). Some discrepancies with previously reported estimates could indicate that the genetic determinism of some genera is affected by the study conditions, either animal dependent (breed, age at sampling, etc.) or related to external conditions (feeding, antibiotic distributions, other management choices, etc.), and would need validation in larger and more diverse conditions.

3.2.5.2. Some genera are genetically correlated with production and FE traits

Obtaining r_g between genera and performance traits lightens the genetic-based interaction between feed efficiency components and gut microbiota composition. About 30% of the studied genera had a significant genetic correlation with a studied trait. However, the number of significant r_g and their magnitudes differed between the five traits. For instance, we could not observe any significant r_g with FCR and ADG, which might be due to the limited power of the analyses. However, this indicates that in our study, the strength of the genetic links between genera and ADG or FCR were lower than with the three other traits.

The negative r_g of the *Streptococcus* genus with RFI and its higher abundance in the HRFI pigs in our study is in agreement with the report of J. Quan et al. (2018). Similarly, our r_g estimate with RFI for the genus *Prevotella_7*, and its lower abundance in LRFI pigs, was consistent with the prevalence of the *Prevotellaceae* family in low versus high FCR pigs reported by J. Quan et al. (2018). Finally, the genus *Desulfovibrio*, that had a negative r_g with RFI and higher abundance in the LRFI pigs, is known as a sulfate-

reducing bacteria that metabolizing sulfites and sulfates of the diet (Gibson, 1990; Kerr et al., 2011). The genus *Desulfovibrio* was also reported with a negative correlation with feed efficiency traits at the phenotypic level in Large White pigs by Bergamaschi, Tiezzi, et al. (2020a). Identifying only three significant r_g with RFI, and none with FCR, seemed very low given the biological assumptions of the key role of gut microbiota on nutrient availability of the host. However, previous studies also showed limited associations between feed efficiency and single microbiota components (H. Yang et al., 2017). Besides biological mechanisms, this could be related to maternal genetic and litter effects involved in the variability of the microbial community that could not be fully accounted for in this analysis. When considering DFI, only the genus *Roseburia* showed significant r_g. The positive r_g with DFI was not in accordance with its higher abundance in the LRFI line, but He et al. (2019) also reported a higher abundance of this *Roseburia* in low FI pigs. Conflict in the line abundances and r_g also suggests that other factors might be driving this genus abundance at the line level (maternal effects, litter effects), that would deserve further analysis.

The higher number of significant r_g between genera and BFT could be partly due to the higher h^2 of BFT, in comparison to the other traits, that could give more power to these estimations. The general composition of backfat in pigs includes water, collagen, and lipids (mainly triacylglycerols) (Wood et al., 1989). Therefore, BFT can be directly affected by the metabolic functions of the microbial composition of the gut. He et al. (2016) have found an association between fatness and OTUs annotated to the genera *Blautia*, *Coprococcus*, and *Ruminococcus* in the cecum samples of pigs. The considerable r_g of the genera *Blautia*, *Coprococcus_3*, *Coprococcus_1*, and *Ruminococcaceae_UCG_008* with BFT in our result is confirming the results of He et al. (2016). Of the 14 genera with significant r_g with BFT, 8 genera (*Blautia*, *Coprococcus_3*, *Syntrophococcus*, *Coprococcus_1*, *Marvinbryantia*, *Dorea*, *Shuttleworthia*, and *Lachnospira*) belonged to the *Lachnospiraceae* family. Biddle et al. (2013) argued that *Lachnospiraceae* and *Ruminococcaceae* families have a role of decomposing substrates from indigestible plant materials of the diet (e.g. cellulose and hemicellulose) in the gut. Compounds resulting from such decomposition would be fermented and converted into the acetate, butyrate, and propionate (short-chain fatty acids - SCFAs) that are absorbable and useable as

energy sources by the host (Biddle et al., 2013). The SCFAs also have essential roles in the composition of the gut environment, maintaining electrolyte balance, and providing energy for host cells as well as gut microbiota (Rios-Covian et al., 2016). Therefore, more availability of SCFAs in the gut environment by the activity of bacteria belonging to the *Lachnospiraceae* and *Ruminococcaceae* families, which have systematic impacts on lipid metabolism and fat storage could justify the chained relationship of these genera with BFT. Given the importance of the BFT as an indicator for carcass payment and reproductive traits of pigs (Roongsitthichai & Tummaruk, 2014), the genetic control of the *Lachnospiraceae* and *Ruminococcaceae* families and the genera belonging to them can have major economic importance in the pig breeding.

3.2.5.3. α-diversity indexes are under genetic control and are related to FE traits

Higher microbial diversity is often considered as an attribute of gut health, as animals with the more diverse microbial community are potentially more capable to better deal with pathogenic microbes (J. M. Fouhse et al., 2016). It has been more generally linked to increased functional redundancies among the microbial community, which can contribute to a more stable metabolic state and better resilience to face larger variability of feeding resources (Moya & Ferrer, 2016). Therefore, microbial diversity is beneficial for the growth performance and productivity of animals (J. M. Fouhse et al., 2016; Hildebrand et al., 2013). This relationship with feed efficiency was confirmed by the negative r_g between the α-diversity metrics and the five traits. Negative correlations imply that selecting animals for improved feed efficiency (lower RFI or FCR) will result in increased intestinal microbial community diversity. In the literature, genetic parameters for α-diversity metrics are rarely reported. Lu et al. (2018), in a study on longitudinal diversity of faecal microbiota in swine, found an h^2 estimate of 0.04 ± 0.04 for the Shannon index at weaning and 0.18 ± 0.08 at week 15 of age. In another study on rumen microbial features in cattle, an h^2 of 0.23 ± 0.09 for the Shannon index and 0.19 ± 0.08 for the Simpson index have been reported (Li et al., 2019). Our estimates of h^2 for both metrics fell into the range of those values. The obtained genetic correlation between the Shannon index and ADG in the present study was lower than -0.53 ± 0.29 reported by Lu et al. (2018). Nevertheless, we have found a stronger r_g between the Shannon index and BFT than their reports (-0.53 ±

0.23 and -0.45 ± 0.25), but given the standard errors in both studies, our estimates are not statistically different from theirs. Given the genetic properties found in our study and the links reported with gut health and immunity, those synthetic descriptors of gut microbiota composition could be promising traits for selection

3.2.5.4. Potential for selection and management in pig production

Our results clearly indicate a genetic basis for part of the gut microbiota composition involved in the variation of feed efficiency (*Streptococcus, Prevotella_7, Desulfovibrio*) and body composition traits (*Lachnospiraceae* family). However, selection to change single microbiota components in order to improve performance traits seems contradictory with the beneficial relationships found between performance traits and microbiota diversity. In that respect, selecting for indicators of microbiota diversity, such as the Shannon index, could be a more generic option. This could also be less dependent on the microbiota specificities due to breeding conditions and sampling characteristics. Indeed, in addition to the genetic, multiple factors can affect the relative abundance of microbiota components and their relationships with traits, including breed and age at sampling (Bergamaschi, Tiezzi, et al., 2020a), breeding environment (Mathilde Le Sciellour et al., 2019), and of course diets (L. M. G. Verschuren et al., 2018). Therefore, more generic indicators of microbiota composition, such as diversity indexes, or mixed models including a microbiability component (Weishaar et al., 2020), might be more relevant for selection. Finally, for some genera (e.g. *Roseburia*) the genetic relationships seemed to be also depend on other factors that could not be accounted for in the present analysis. Deciphering the role of these different factors (genetics, litter and maternal for instance) would clarify the potential for use of these microbiota components to orientate pig performances via different levers of management, including the use of pro and pre-biotics, as proposed by Maltecca et al. (2020).

3.2.6. Conclusion

Our results showed substantial effects of genetics on the variability of gut genera community and their relationship with the feed efficiency in pigs. Both analyses of line effect and genetic correlations with production traits revealed a substantial genetic basis

for the links between feed efficiency traits and genera and individual diversity of the gut microbial community. The higher diversity in more feed efficient pigs might be related to better gut health and resilience to feed changes. Genera annotated to the *Lachnospiraceae* family had more significant correlations with the studied traits than genera from other families. Functional analyses will be needed to validate the underlying mechanisms. The robustness of these findings requires further validations in different breeding conditions. However, they offer promising perspectives for selection for feed efficiency using gut microbiome composition in pigs.

Acknowledgments

The authors thank financial support of the French National Research Agency via the Micro-Feed project, under grant ANR-16-CE20-0003. The authors would like to thank also the experimental farm staff for data collection and breeding of the animals.

Competing interests

The authors declare that they have no competing interests.

3.2.7. References

Achard, C. S., Dupouy, V., Cauquil, L., Arpaillange, N., Bousquet-Melou, A., Floc'h, N. L., & Zemb, O. (2020). Early Inoculation of Microbial Suspension in Suckling Piglets Affects the Transmission of Maternal Microbiota and the Associated Antibiotic Resistance Genes. *Microorganisms*, *8*, 1576. https://doi.org/10.3390/microorganisms8101576

Aliakbari, A., Delpuech, E., Labrune, Y., Riquet, J., & Gilbert, H. (2020). The impact of training on data from genetically-related lines on the accuracy of genomic predictions for feed efficiency traits in pigs. *Genetics Selection Evolution*, *52*, 57. https://doi.org/10.1186/s12711-020-00576-0

Bergamaschi, M., Maltecca, C., Schillebeeckx, C., McNulty, N. P., Schwab, C., Shull, C., Fix, J., & Tiezzi, F. (2020). Heritability and genome-wide association of swine gut microbiome features with growth and fatness parameters. *Sci Rep*, *10*, 10134. https://doi.org/10.1038/s41598-020-66791-3

Bergamaschi, M., Tiezzi, F., Howard, J., Huang, Y. J., Gray, K. A., Schillebeeckx, C., McNulty, N. P., & Maltecca, C. (2020). Gut microbiome composition differences among breeds impact feed efficiency in swine. *Microbiome*, *8*, 110. https://doi.org/10.1186/s40168-020-00888-9

Biddle, A., Stewart, L., Blanchard, J., & Leschine, S. (2013). Untangling the genetic basis of fibrolytic specialization by Lachnospiraceae and Ruminococcaceae in diverse gut communities. *Diversity*, *5*, 627-640.

Brenner, S., Miller, J. H., & Broughton, W. (2002). *Encyclopedia of genetics*.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*, *11*, 2639-2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581-583. https://doi.org/10.1038/nmeth.3869

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S., & Bennewitz, J. (2017). Host Genome Influence on Gut Microbial Composition and Microbial

Prediction of Complex Traits in Pigs. *Genetics*, *206*, 1637-1644. https://doi.org/10.1534/genetics.117.200782

Chatelet, A., Gondret, F., Merlot, E., Gilbert, H., Friggens, N. C., & Le Floc'h, N. (2018). Impact of hygiene of housing conditions on performance and health of two pig genetic lines divergent for residual feed intake. *Animal*, *12*, 350-358. https://doi.org/10.1017/S1751731117001379

Chen, C., Huang, X., Fang, S., Yang, H., He, M., Zhao, Y., & Huang, L. (2018). Contribution of Host Genetics to the Variation of Microbial Composition of Cecum Lumen and Feces in Pigs. *Front Microbiol*, *9*, 2626. https://doi.org/10.3389/fmicb.2018.02626

Dowarah, R., Verma, A. K., & Agarwal, N. (2017). The use of Lactobacillus as an alternative of antibiotic growth promoters in pigs: A review. *Animal Nutrition*, *3*, 1-6. https://doi.org/10.1016/j.aninu.2016.11.002

du Toit, M., Huch, M., Cho, G. S., & Franz, C. (2014). The genus Streptococcus. *Lactic Acid Bacteria*, 457-505. https://doi.org/doi.org/10.1002/9781118655252.ch28

Fouhse, J. M., Zijlstra, R. T., & Willing, B. P. (2016). The role of gut microbiota in the health and disease of pigs. *Animal Frontiers*, *6*, 30-36. https://doi.org/10.2527/af.2016-0031

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., & Graves, S. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing*.

Giang, H. H., Vieti, T. Q., Ogle, B., & Lindberg, J. E. (2011). Effects of Supplementation of Probiotics on the Performance, Nutrient Digestibility and Faecal Microflora in Growing-finishing Pigs. *Asian-australasian journal of animal sciences*, *24*, 655-661. https://doi.org/10.5713/ajas.2011.10238

Gibson, G. R. (1990). Physiology and ecology of the sulphate-reducing bacteria. *Journal of Applied Bacteriology*, *69*, 769-797. https://doi.org/10.1111/j.1365-2672.1990.tb01575.x

Gilbert, H., Bidanel, J. P., Gruand, J., Caritez, J. C., Billon, Y., Guillouet, P., Lagant, H., Noblet, J., & Sellier, P. (2007). Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality

traits. *Journal of Animal Science*, *85*, 3182-3188. https://doi.org/10.2527/jas.2006-590

Gilbert, H., Billon, Y., Brossard, L., Faure, J., Gatellier, P., Gondret, F., Labussiere, E., Lebret, B., Lefaucheur, L., Le Floch, N., Louveau, I., Merlot, E., Meunier-Salaun, M. C., Montagne, L., Mormede, P., Renaudeau, D., Riquet, J., Rogel-Gaillard, C., van Milgen, J., Vincent, A., & Noblet, J. (2017). Review: divergent selection for residual feed intake in the growing pig. *Animal*, *11*, 1427-1439. https://doi.org/10.1017/S175173111600286X

He, B. B., Li, T. T., Wang, W., Gao, H., Bai, Y., Zhang, S., Zang, J. J., Li, D. F., & Wang, J. J. (2019). Metabolic characteristics and nutrient utilization in high-feed-efficiency pigs selected using different feed conversion ratio models. *Science China-Life Sciences*, *62*, 959-970. https://doi.org/10.1007/s11427-018-9372-6

He, M., Fang, S., Huang, X., Zhao, Y., Ke, S., Yang, H., Li, Z., Gao, J., Chen, C., & Huang, L. (2016). Evaluating the Contribution of Gut Microbiota to the Variation of Porcine Fatness with the Cecum and Fecal Samples. *Front Microbiol*, *7*, 2108. https://doi.org/10.3389/fmicb.2016.02108

Hildebrand, F., Nguyen, T. L., Brinkman, B., Yunta, R. G., Cauwe, B., Vandenabeele, P., Liston, A., & Raes, J. (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol*, *14*, R4. https://doi.org/10.1186/gb-2013-14-1-r4

Hill, W. G. (1972). *Estimation of genetic change: General theory and design of control populations*.

Ingala, M. R., Simmons, N. B., Wultsch, C., Krampis, K., Speer, K. A., & Perkins, S. L. (2018). Comparing microbiome sampling methods in a wild mammal: fecal and intestinal samples record different signals of host ecology, evolution. *Frontiers in microbiology*, *9*, 803.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*, 149-176. https://www.ncbi.nlm.nih.gov/pubmed/18132090

Kerr, B. J., Weber, T. E., Ziemer, C. J., Spence, C., Cotta, M. A., & Whitehead, T. R. (2011). Effect of dietary inorganic sulfur level on growth performance, fecal composition, and measures of inflammation and sulfate-reducing bacteria in the

intestine of growing pigs. *Journal of Animal Science*, *89*, 426-437. https://doi.org/10.2527/jas.2010-3228

Kim, E. Y., Kim, Y. H., Rhee, M. H., Song, J. C., Lee, K. W., Kim, K. S., Lee, S. P., Lee, I. S., & Park, S. C. (2007). Selection of Lactobacillus sp. PSC101 that produces active dietary enzymes such as amylase, lipase, phytase and protease in pigs. *Journal of General and Applied Microbiology*, *53*, 111-117. https://doi.org/10.2323/jgam.53.111

Le Floc'h, N., Knudsen, C., Gidenne, T., Montagne, L., Merlot, E., & Zemb, O. (2014). Impact of feed restriction on health, digestion and faecal microbiota of growing pigs housed in good or poor hygiene conditions. *Animal*, *8*, 1632-1642. https://doi.org/10.1017/S1751731114001608

Le Sciellour, M., Zemb, O., Hochu, I., Riquet, J., Gilbert, H., Giorgi, M., Billon, Y., Gourdine, J.-L., & Renaudeau, D. (2019). Effect of chronic and acute heat challenges on fecal microbiota composition, production, and thermoregulation traits in growing pigs1,2. *Journal of Animal Science*, *97*, 3845-3858. https://doi.org/10.1093/jas/skz222

Li, F., Li, C., Chen, Y., Liu, J., Zhang, C., Irving, B., Fitzsimmons, C., Plastow, G., & Guan, L. L. (2019). Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome*, *7*, 92. https://doi.org/10.1186/s40168-019-0699-1

Lu, D., Tiezzi, F., Schillebeeckx, C., McNulty, N. P., Schwab, C., Shull, C., & Maltecca, C. (2018). Host contributes to longitudinal diversity of fecal microbiota in swine selected for lean growth. *Microbiome*, *6*, 4. https://doi.org/10.1186/s40168-017-0384-1

Magoc, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957-2963. https://doi.org/10.1093/bioinformatics/btr507

Maltecca, C., Bergamaschi, M., & Tiezzi, F. (2020). The interaction between microbiome and pig efficiency: A review. *Journal of Animal Breeding and Genetics*, *137*, 4-13. https://doi.org/10.1111/jbg.12443

McCormack, U. M., Curiao, T., Buzoianu, S. G., Prieto, M. L., Ryan, T., Varley, P., Crispie, F., Magowan, E., Metzler-Zebeli, B. U., Berry, D., O'Sullivan, O., Cotter, P. D., Gardiner, G. E., & Lawlor, P. G. (2017). Exploring a Possible Link between the Intestinal Microbiota and Feed Efficiency in Pigs. *Applied and Environment Microbiology*, *83*. https://doi.org/10.1128/AEM.00380-17

Moya, A., & Ferrer, M. (2016). Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends in Microbiology*, *24*, 402-413. https://doi.org/10.1016/j.tim.2016.02.002

Muscogiuri, G., Cantone, E., Cassarano, S., Tuccinardi, D., Barrea, L., Savastano, S., Colao, A., on behalf of the Obesity Programs of nutrition, E. R., & Assessment, g. (2019). Gut microbiota: a new path to treat obesity. *Int J Obes Suppl*, *9*, 10-19. https://doi.org/10.1038/s41367-019-0011-7

Niu, Q., Li, P. H., Hao, S. S., Kim, S. W., Du, T. R., Hua, J. D., & Huang, R. H. (2019). Characteristics of Gut Microbiota in Sows and Their Relationship with Apparent Nutrient Digestibility. *International journal of molecular sciences*, *20*, 870. https://doi.org/10.3390/ijms20040870

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2013). Package 'vegan'. *Community ecology package, version*, *2*, 1-295.

Ottosen, M., Mackenzie, S. G., Wallace, M., & Kyriazakis, I. (2020). A method to estimate the environmental impacts from genetic change in pig production systems. *International Journal of Life Cycle Assessment*, *25*, 523-537. https://doi.org/10.1007/s11367-019-01686-8

Plummer, E., Bulach, D., Carter, G., & Albert, M. J. (2020). Gut microbiome of native Arab Kuwaitis. *Gut Pathog*, *12*, 10. https://doi.org/10.1186/s13099-020-00351-y

Quan, J., Cai, G., Ye, J., Yang, M., Ding, R., Wang, X., Zheng, E., Fu, D., Li, S., Zhou, S., Liu, D., Yang, J., & Wu, Z. (2018). A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios. *Sci Rep*, *8*, 4536. https://doi.org/10.1038/s41598-018-22692-0

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project:

improved data processing and web-based tools. *Nucleic Acids Research*, *41*, 590-596. https://doi.org/10.1093/nar/gks1219

Rios-Covian, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., de los Reyes-Gavilan, C. G., & Salazar, N. (2016). Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Frontiers in microbiology*, *7*, 185. https://doi.org/10.3389/fmicb.2016.00185

Roongsitthichai, A., & Tummaruk, P. (2014). Importance of backfat thickness to reproductive performance in female pigs. *The Thai Journal of Veterinary Medicine*, *44*, 171-178.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*, 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shon, K. S., Hong, J. W., Kwon, O. S., Min, B. J., Lee, W. B., Kim, I. H., Park, Y. H., & Lee, I. S. (2005). Effects of Lactobacillus reuteri-based direct-fed microbial supplementation for growing-finishing pigs. *Asian-australasian journal of animal sciences*, *18*, 370-374. https://doi.org/10.5713/ajas.2005.370

Si, J. L., Feng, L. L., Gao, J. Y., Huang, Y., Zhang, G. J., Mo, J. Y., Zhu, S. R., Qi, W. J., Liang, J., & Lan, G. Q. (2020). Evaluating the association between feed efficiency and the fecal microbiota of early-life Duroc pigs using 16S rRNA sequencing. *AMB Express*, *10*, 115. https://doi.org/10.1186/s13568-020-01050-2

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*, 688. https://doi.org/10.1038/163688a0

Valeriano, V. D., Balolong, M. P., & Kang, D. K. (2017). Probiotic roles of Lactobacillus sp. in swine: insights from gut microbiota. *Journal of Applied Microbiology*, *122*, 554-567. https://doi.org/10.1111/jam.13364

Verschuren, L. M. G., Calus, M. P. L., Jansman, A. J. M., Bergsma, R., Knol, E. F., Gilbert, H., & Zemb, O. (2018). Fecal microbial composition associated with variation in feed efficiency in pigs depends on diet and sex. *Journal of Animal Science*, *96*, 1405-1418. https://doi.org/10.1093/jas/sky060

Wang, X., Jordan, I. K., & Mayer, L. W. (2015). A phylogenetic perspective on molecular epidemiology. In *Molecular Medical Microbiology* (pp. 517-536). Academic Press.

Weishaar, R., Wellmann, R., Camarinha-Silva, A., Rodehutscord, M., & Bennewitz, J. (2020). Selecting the hologenome to breed for an improved feed efficiency in pigs—A novel selection index. *Journal of Animal Breeding and Genetics*, *137*, 14-22. https://doi.org/10.1111/jbg.12447

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, *30*, 280-338. https://doi.org/10.2307/1943563

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev*, *51*, 221-271. https://www.ncbi.nlm.nih.gov/pubmed/2439888

Wood, J. D., Enser, M., Whittington, F. M., Moncrieff, C. B., & Kempster, A. J. (1989). Backfat Composition in Pigs - Differences between Fat Thickness Groups and Sexes. *Livestock production science*, *22*, 351-362. https://doi.org/10.1016/0301-6226(89)90066-3

Yang, H., Huang, X., Fang, S., He, M., Zhao, Y., Wu, Z., Yang, M., Zhang, Z., Chen, C., & Huang, L. (2017). Unraveling the Fecal Microbiota and Metagenomic Functional Capacity Associated with Feed Efficiency in Pigs. *Front Microbiol*, *8*, 1555. https://doi.org/10.3389/fmicb.2017.01555

Yu, H. F., Wang, A. N., Li, X. J., & Qiao, S. Y. (2008). Effect of viable Lactobacillus fermentum on the growth performance, nutrient digestibility and immunity of weaned pigs. *Journal of Animal and Feed Sciences*, *17*, 61-69. https://doi.org/10.22358/jafs/66470/2008

3.2.8. Appendix 3.1

Table S1 P-values of the fixed effects† tested with linear models on α-diversity indexes and the 75 genera

	BW at Test	CG	Sex	Herd	Pen size
α-diversity index					
Shannon	0.2109	0.0004	0.2340	0.3416	<0.0001
Simpson	0.2197	0.0343	0.6425	0.3446	0.0006
Genus					
Clostridium_sensu_stricto_1	0.0368	<0.0001	0.8528	0.1755	0.0056
Prevotella_1	0.7835	<0.0001	0.2644	0.0579	0.0018
Blautia	0.0341	<0.0001	0.0202	0.4815	0.0005
Prevotellaceae_NK3B31_group	0.4067	<0.0001	0.0901	0.2988	0.012
Lachnospiraceae_NK3A20_group	0.599	0.0004	0.3593	0.9592	0.002
Ruminococcaceae_UCG-008	0.001	<0.0001	0.2229	0.8803	0.6962
Lachnospiraceae_ND3007_group	0.0217	0.0026	0.1102	0.8725	0.2676
Coprococcus_3	0.1679	<0.0001	0.166	0.8281	0.9912
Butyricicoccus	0.0211	0.0002	0.0099	0.5954	0.001
Terrisporobacter	0.0309	<0.0001	0.6791	0.3966	0.0001
Syntrophococcus	0.321	<0.0001	0.5874	0.28	0.1499
Faecalibacterium	0.0247	<0.0001	0.0003	0.2983	<0.0001
Coprococcus_1	0.6799	<0.0001	0.4349	0.8021	0.5994
Marvinbryantia	0.4238	<0.0001	0.086	0.3768	0.6486
Mitsuokella	0.2467	<0.0001	0.0811	0.0459	0.113
NA_Family_XIII	0.3531	<0.0001	0.6175	0.9327	0.1163
Prevotella_7	0.8175	<0.0001	0.3251	0.8211	0.3179
Prevotellaceae_UCG-003	0.9624	<0.0001	0.2044	0.1044	0.0659
Romboutsia	0.4618	<0.0001	0.8058	0.0629	<0.0001
Fusicatenibacter	0.0106	0.0006	0.0107	0.1828	0.0001
Campylobacter	0.4882	<0.0001	0.0164	0.9164	0.0258
Olsenella	0.6392	<0.0001	0.3613	0.1878	0.2266
Oscillospira	0.6924	<0.0001	0.022	0.232	0.0007
Lactobacillus	0.911	<0.0001	0.6363	0.6634	<0.0001
Roseburia	<0.0001	<0.0001	0.109	0.0453	0.0312
Succinivibrionaceae_UCG-001	0.7886	<0.0001	0.6601	0.2291	0.0832
NA_Muribaculaceae	0.3327	<0.0001	0.6308	0.3702	<0.0001

Dorea	0.0015	0.0229	0.3121	0.7921	0.0461
Subdoligranulum	0.0446	<0.0001	0.0034	0.1668	0.0021
Alloprevotella	0.4935	<0.0001	0.606	0.4329	0.0065
Ruminococcaceae_UCG-014	0.2208	<0.0001	0.0059	0.0882	0.0002
Dialister	0.7177	<0.0001	0.0477	0.0896	0.0725
Shuttleworthia	0.7231	<0.0001	0.4121	0.9095	0.664
Streptococcus	0.7774	<0.0001	0.0003	0.5731	0.0002
NA_Prevotellaceae	0.6695	0.0099	0.0223	0.26	0.0003
Rikenellaceae_RC9_gut_group	0.5099	<0.0001	0.7212	0.4198	<0.0001
Lachnospiraceae_NK4A136_group	0.7686	<0.0001	0.0497	0.3305	0.1102
Desulfovibrio	0.4229	<0.0001	0.652	0.017	0.0004
Lachnospiraceae_UCG-001	0.0103	<0.0001	0.58	0.2758	0.9719
Ruminococcus_2	0.0342	<0.0001	0.0003	0.5176	0.0084
NA_Ruminococcaceae	0.2133	<0.0001	0.6919	0.5629	<0.0001
Treponema_2	0.8724	<0.0001	0.1673	0.1082	0.004
Fournierella	0.0172	0.0027	0.19	0.3109	0.0105
Prevotella_2	0.0452	<0.0001	0.1835	0.511	0.0637
Agathobacter	0.0014	<0.0001	0.0044	0.8655	0.0003
Lachnospira	0.1553	<0.0001	0.0036	0.0591	<0.0001
Ruminococcaceae_UCG-005	0.9505	<0.0001	0.3494	0.3662	0.0001
Lachnospiraceae_UCG-004	0.2141	0.001	0.004	0.3512	0.0003
Oribacterium	0.1649	<0.0001	<0.0001	0.2102	<0.0001
Ruminiclostridium_5	0.2015	0.0001	0.398	0.6998	<0.0001
Family_XIII_AD3011_group	0.2655	<0.0001	0.3105	0.8536	0.0001
Christensenellaceae_R-7_group	0.2423	<0.0001	0.0837	0.9878	0.0057
Lachnospiraceae_FCS020_group	0.425	<0.0001	0.0378	0.7094	0.2284
NA_NA_Bradymonadales	0.671	<0.0001	0.4284	0.4882	0.0083
Family_XIII_UCG-001	0.541	<0.0001	0.005	0.6487	0.2152
Mogibacterium	0.8338	<0.0001	0.6873	0.6786	0.0148
Succinivibrio	0.3089	0.0003	0.1552	0.9261	0.2504
Ruminococcaceae_UCG-013	0.26	<0.0001	0.8237	0.0299	<0.0001
Intestinimonas	0.6018	0.1622	0.9774	0.6654	0.2347
Turicibacter	0.2598	<0.0001	0.8787	0.0582	<0.0001
Intestinibacter	0.0947	0.0001	0.8015	0.9285	0.6409
Ruminococcaceae_UCG-002	0.0772	<0.0001	0.967	0.7606	<0.0001
NA_Lachnospiraceae	0.299	<0.0001	0.0009	0.2538	0.0001
Ruminococcaceae_UCG-010	0.7855	<0.0001	0.1223	0.3466	0.0011

Prevotellaceae_UCG-001	0.3501	<0.0001	0.1161	0.9331	0.0002
Prevotella_9	0.0006	<0.0001	0.0142	0.9962	0.0002
NA_NA_Bacteroidales	0.4936	<0.0001	<0.0001	0.8061	0.0007
Coprococcus_2	0.0083	<0.0001	0.0002	0.0376	<0.0001
Peptococcus	0.0353	0.0048	0.161	0.3722	0.5593
Ruminococcus_1	0.7002	<0.0001	0.4908	0.5111	0.276
NA_Eggerthellaceae	0.5374	0.0007	0.1811	0.4583	0.0245
Ruminiclostridium_9	0.8487	0.0426	0.0031	0.6418	0.012
Lachnoclostridium	0.6725	<0.0001	0.1285	0.1093	0.0614
Ruminococcaceae_NK4A214_group	0.1475	<0.0001	0.9606	0.1203	<0.0001
Parabacteroides	0.2206	<0.0001	0.7058	0.0308	0.3969

[†]BW = body weight, CG = contemporary group. Coloured cells show significant effects (P-value < 0.05

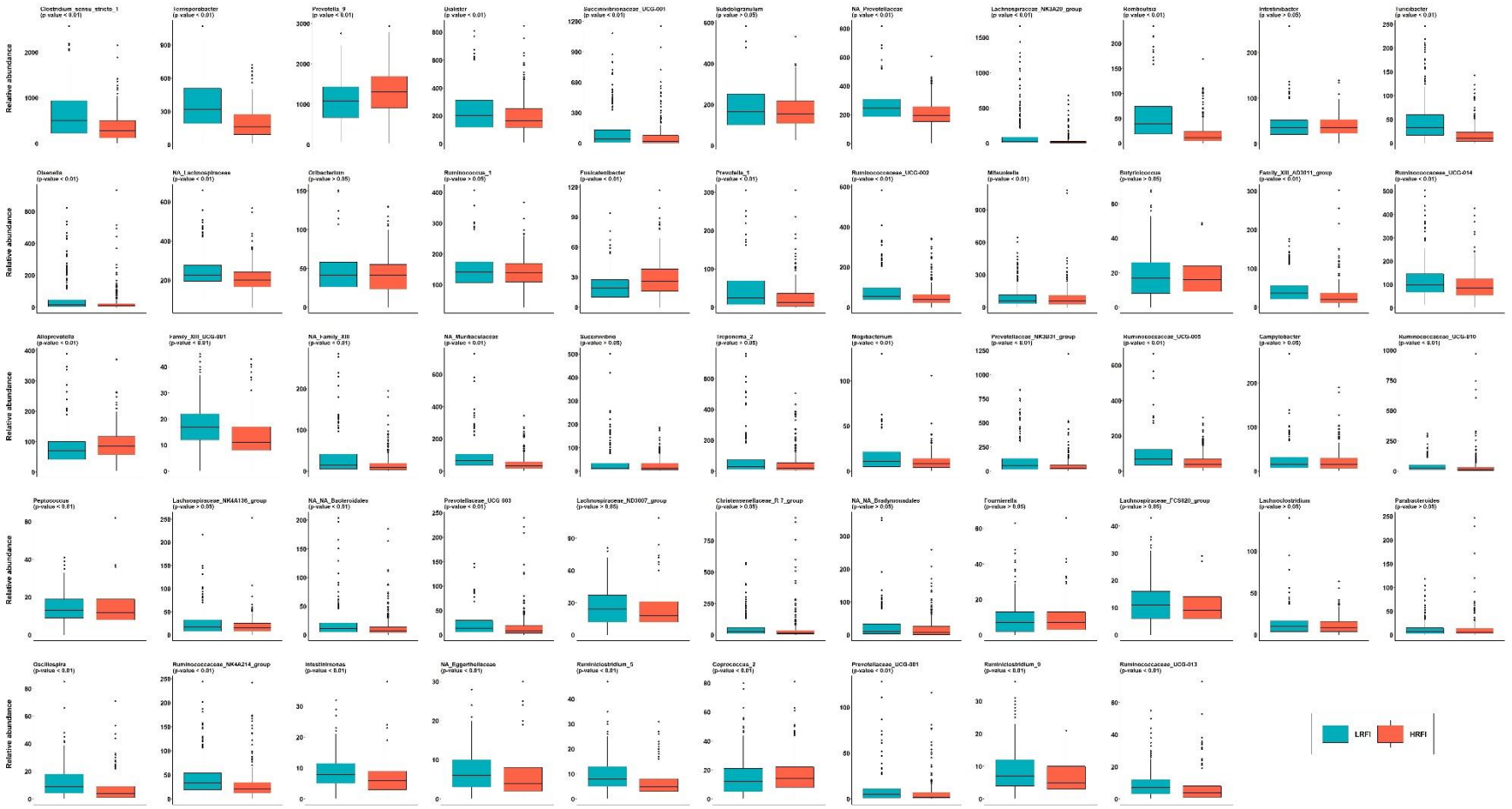


Figure S1. Box plots showing genera abundances between the LRFI and HRFI lines for the other 53 genera

3.2.9. Appendix 3.2

A short paper submitted to the EAAP-2020 fellowship

71. Annual meeting of the European Federation of Animal Science (EAAP), Virtual Meeting, Dec, 2020

Genetic relationships between feed efficiency and fecal microbiome in pig lines

Amir ALIAKBARI[1], Olivier ZEMB[1], Yvon BILLON[2], Céline BARILLY[1], Ingrid AHN[1], Juliette RIQUET[1] and Hélène GILBERT[1]

[1]GenPhySE, Université de Toulouse, INRAE, F-31326 Castanet-Tolosan, France
[2]GenESI, INRAE, F-17700 Surgères, France

Summary

Recent advances in bioinformatics and sequencing technologies have made it possible to obtain individual microbiome information for human, animals and plants. In pigs, as in humans, gut microbiota is an important contributor to the nutrient availability at the gut level. In the present study we aimed to quantify the genetic relationships between two main feed efficiency traits, feed conversion ratio (FCR) and residual feed intake (RFI), and fecal microbial composition in two experimental pig lines divergently selected for RFI (HRFI and LRFI lines). Multivariate linear mixed models of OTUs relative abundancies and performance traits provided heritability (h^2) and additive genetic correlation (r_A) for all traits. Fecal samples were collected at 15 weeks of age in 604 pigs from the G9 and G10 generations of the RFI lines, and about 4000 FCR and RFI records were available for all generations. From sequencing of the V3-V4 regions of the16S rRNA gene, a total of 6792 Operational Taxonomic Units (OTU) were identified in the samples. The 137 OTUs with less than 20% zero abundancies were kept for genetic analyses after log-transformation.

A total of 65 OTUs showed a rh^2 different from zero ($P<0.05$), with estimates ranging from 0.13 ± 0.07 to 0.52 ± 0.12. In total, OTUs with significant h^2 were annotated to 13 families and 34 genera. Among those 65 OTUs with genetic background, 10 OTUs had a genetic correlation with FCR different from zero, and 14 OTUs had a significant genetic correlation with RFI. The OTUs with significant correlations with FCR belonged to six families. The OTUs with significant r_A with RFI belonged to four families. Only one OTU, belonging to the Prevotella_9 genus from the Prevotellaceae family, had commonly significant correlation with both traits. Our results showed that some OTU abundancies have a genetic background and significant genetic correlation with feed efficiency traits. These results beside the host genetic effect could deserve more consideration in breeding programs to improve the feed efficiency in pigs.

Key words: divergent lines, feed efficiency, genetics, gut microbiota, swine

Introduction

Recent advances in bioinformatics and sequencing technologies have made it possible to obtain individual microbiome information for human, animals and plants. In pig breeding, feed efficiency (FE), because of its contributions to economic and environmental pillars of the production, has a high impact on the sustainability of this industry. The gut microbial composition, besides the effects on physiological health of the pigs, has a main role in nutrient digestibility (J. Fouhse et al., 2016; Qing Niu et al., 2019). Therefore, measuring the magnitude of the genetic control on gut microbiota information and its genetic correlation with feed efficiency and production traits can provide insights into potential benefits from this new information in animal breeding. In the present study we aimed to quantify the genetic relationships between two main FE traits, feed conversion ratio (FCR) and residual feed intake (RFI), and fecal microbial composition in two experimental pig lines developed by divergent selection for RFI (HRFI and LRFI lines).

Methods

Population and dataset

The data were collected from a selection experiment conducted at INRAE (UE GenESI, Surgères, France) in the French Large White pig breed. The two lines were established by 10 generations of divergent selection for RFI (based on an RFI index) from 2000 to 2015 (Gilbert et al., 2007). The initial matings (F0) were conducted by artificial insemination between 30 boars and 30 gilts. From resulted G0 population, 116 boar were tested for RFI as candidates for selection. Among tested animals six founder boars for the low RFI (LRFI) line and six founder boars for high RFI (HRFI) line were selected. The two lines were then initiated by mating these boars to about 35 G0 gilts per line. From G1 to G10, six boars were selected from 96 candidates in each generation. At least one additional parity was produced in each generation to evaluate the correlated responses to selection on production traits on both females and castrated males (response animals) leading to a total of 3802 records for RFI and 4282 records for FCR.

Microbial DNA extraction and 16S rRNA gene sequencing

Fecal samples of 604 animals from G9 and G10 pigs of both lines were sampled at 15 weeks of age and stored at -80°C until being used for ribosomal 16S DNA gene sequencing and analysis. Microbial profiling was done by amplification of the V3-V4 region of the 16S rRNA gene extracted from purified DNA. Amplification was done in 30 cycles with annealing temperature of 65°C. The purified PCR products were sequenced using Illumina MiSeq cartridge according to the manufacturer instructions at the GetPlaGe platform. After high-throughput sequencing, filtered and trimmed sequences of high quality were clustered into OTUs based on 97% identity of the reads with DADA2 (Callahan et al., 2016). The clustering step was followed by species annotation and indication of OTU phylogeny based on Silva Dataset v132. The final OTU table contained 6792 OTUs for 604 individuals. Rarefaction (with sample size equal to 9000) was applied to the OTU table to correct for differences in sampling efforts (McMurdie & Holmes, 2014). Finally, the log-transformed table of OTU counts was filtered for a maximum of 20 % zero values per OTU. After the filtration, 137 OTUs remained for the downstream analyses.

Model and analyses

Variance components and genetic parameters were estimated for OTUs using the following animal mixed model in bivariate (selection index and one OTU) and three-variate (selection index, one OTU and FCR or RFI) approaches: $\mathbf{y} = \mathbf{Xb} + \mathbf{Z_1a} + \mathbf{Z_2l} + \mathbf{e}$, where **y** is the vector of observations, **b** is the vector of fixed effects, **a** is the vector of additive genetic effects, **l** is the vector of litter effects and **e** is the vector of random residuals. **X**, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices for **b**, **a** and **l**. The distributions assumed for the random terms were $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G_0} \otimes \mathbf{A})$, $\mathbf{l} \sim N(\mathbf{0}, \mathbf{R_l} \otimes \mathbf{I})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R_e} \otimes \mathbf{I})$, where **A** was the pedigree relationship matrix, $\mathbf{G_0}$ was (co)variance matrix of direct additive genetic effect, and $\mathbf{R_l}$ and $\mathbf{R_e}$ were (co)variance matrices of litter effect and residual effect, respectively. **I** denoted the identity matrix. The **A** included 10 generations of pedigree information plus ancestors, and contained 7293 animals. The analyses were performed using AIREMLF90 software (Misztal et al., 2018) for BLUP method. The litter effect was significant ($P < 0.05$) for 11 OTUs only.

Significance tests

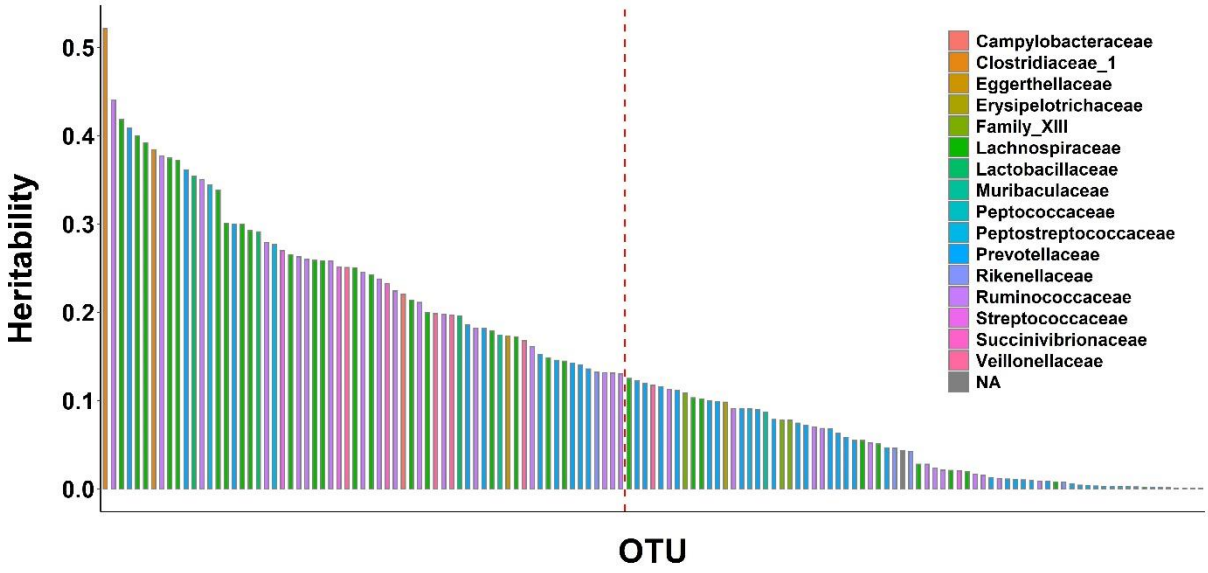
A significance threshold for the estimated heritabilities was estimated to 0.125. This threshold was obtained after running 10000 univariate analyses under the null hypothesis of no genetic control on the OTU, for two OTUs (OTUs with lowest and highest h^2 with the bivariate analyses). The null hypothesis was obtained by shuffling the OTUs counts. The minimum value of the 5 % highest estimated h^2 was considered as the threshold to decide that an OTU was heritable. Thereafter, the three-variate analyses were conducted for OTUs with heritabilities significantly different from zero. The deviation from zero of the additive genetic correlations between OTUs and the two studied traits were tested using the Z-test.

Results and discussion

The estimated heritabilities of OTUs with the bivariate analyses are given in Figure 1. From 137 OTUs, 65 OTUs showed a significant h^2, ranging from 0.13 ± 0.07 to 0.52 ± 0.12. In total, OTUs with significant h^2 were annotated to 13 families and 34 genera, out of an initial distribution among 88 families and 260 genera. The lowest h^2 was observed

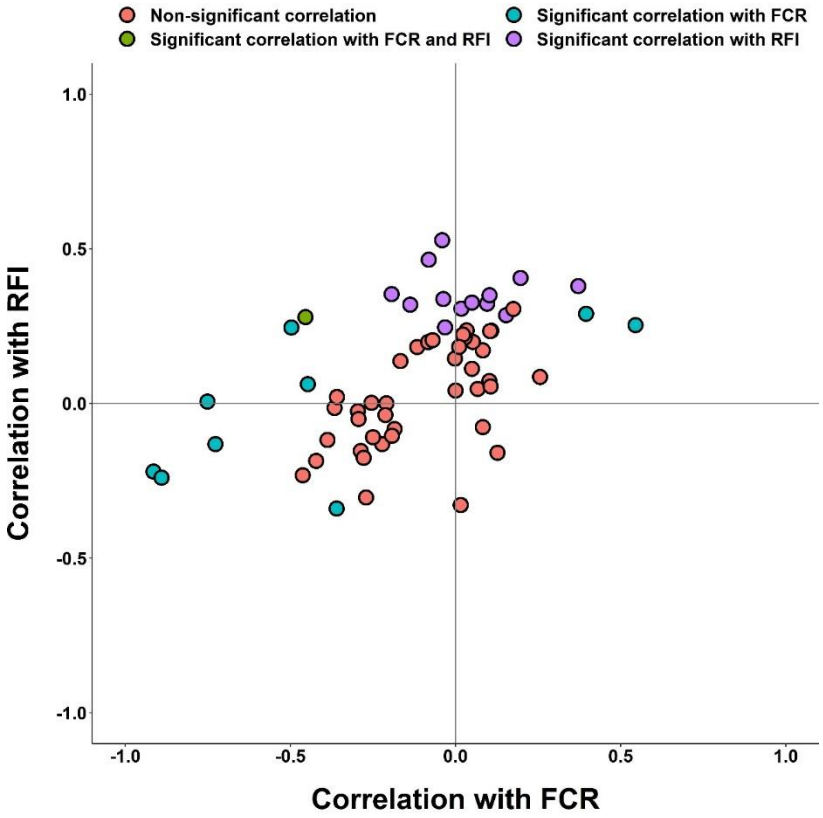
for an OTU annotated to the Ruminococcaceae family and the highest h^2 was annotated to the Clostridiaceae_1 family.

Figure 1. Heritability of OTUs by family designation obtained using bivariate linear mixed models



Few studies have implemented mixed models equations to assess the variance components and genetic relationships between gut microbiota information and feed efficiency in pigs. A study on pigs using colon digesta samples, a h^2 of 0.33 for the Blautia and 0.34 for the Lactobacillus genera have been reported (Camarinha-Silva et al., 2017). Even though that in the present study we have not reported results of analyses at genus level, we observed a similar average h^2 of OTUs annotated to these genera. The genetic correlations of OTUs with FCR and RFI are shown in Figure 2. From the 65 OTUs with significant genetic background, 10 OTUs had genetic correlations different from zero with FCR, ranging from -0.91 ± 0.04 to 0.54 ± 0.16, and 14 OTUs had significant genetic correlations with RFI, ranging from 0.25 ± 0.12 to 0.53 ± 0.11.

Figure 2. Genetic correlations between OTUs and FCR and RFI



The OTUs with significant correlation with FCR belonged to the Lactobacillaceae, Rikenellaceae, Lachnospiraceae, Ruminococcaceae, Prevotellaceae and Peptostreptococcaceae families and those with significant correlation with RFI belonged to the Lachnospiraceae, Prevotellaceae, Streptococcaceae and Ruminococcaceae families. At the genus level, two OTUs annotated to Lactobacillus and Prevotella_2 had positive genetic correlations with FCR and the other 8 OTUs annotated to Rikenellaceae_RC9_gut_group, Subdoligranulum, Prevotella_9, Romboutsia, Prevotella_7 and Agathobacter had negative genetic correlations with FCR. Only one OTU belonging to the Prevotella_9 genus from the Prevotellaceae family had commonly significant correlation with both traits. The other seven genera related to RFI included Blautia, Prevotellaceae_NK3B31_group, Streptococcus, Faecalibacterium, Subdoligranulum, Fusicatenibacter, and Dorea. Altogether, we observed 10 OTUs that had opposite direction of significant genetic correlation with FCR and RFI, which can be

considered as a set of OTUs that are affecting the feed efficiency. At the phenotypic level, Hui Yang et al. (2017) showed positive associations between Lachnospiraceae and Ruminococcaceae and porcine feed efficiency. In our study, both families had negative genetic correlations with feed efficiency.

In conclusion, our results showed that the abundance of some OTUs has a genetic background and can be inherited from one generation to the next. In addition, we have seen interesting genetic correlations between some fecal microbiota information and feed efficiency traits that will need to be confirmed in external datasets. Altogether, having a prior knowledge about the genetic variance components of OTUs abundances that are related to some key microbial genera and families could provide insights to a joint selection of pigs based on fecal microbiota information and performances.

Acknowledgements

The authors would like to thank the experimental farm staff for data collection and breeding of the animals, and the French National Research Agency via the MicroFeed project for financial support.

References

1. Fouhse J, Zijlstra R, Willing B. The role of gut microbiota in the health and disease of pigs. Animal Frontiers. 2016;6:30-6.
2. Niu Q, Li P, Hao S, Kim SW, Du T, Hua J, et al. Characteristics of gut microbiota in sows and their relationship with apparent nutrient digestibility. International journal of molecular sciences. 2019;20:870.
3. Gilbert H, Bidanel J-P, Gruand J, Caritez J-C, Billon Y, Guillouet P, et al. Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. J Anim Sci. 2007;85:3182-8.
4. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581.
5. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comp Biol. 2014;10:e1003531.

6. Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A, et al. Manual for BLUPF90 family of programs. Vol. 2018. 2018.

7. Camarinha-Silva A, Maushammer M, Wellmann R, Vital M, Preuss S, Bennewitz J. Host genome influence on gut microbial composition and microbial prediction of complex traits in pigs. Genetics. 2017;206:1637-44.

8. Yang H, Huang X, Fang S, He M, Zhao Y, Wu Z, et al. Unraveling the fecal microbiota and metagenomic functional capacity associated with feed efficiency in pigs. Frontiers in microbiology. 2017;8:1555.

4.

How microbiota contributes to the variability of the production traits

4.1. Introduction

Recent advances in obtaining microbiota information enable surveying the interplay between complex traits and the microbial community of the gastrointestinal tract (GIT). After testing the genetic background of microbiota genera and α-diversity indexes, and their genetic relationships with feed efficiency traits in chapter 3, the objective of the present chapter was to decipher how microbiota contributes to the variability of the production traits. This was examined in two steps:

1- By investigating the contribution of faecal microbial variants to the variance of the five studied traits including ADG, BFT, DFI, FCR, and RFI.

2- By performing microbiome-wide association studies (MWAS) based on two methods, single-OTU regressions and back solving of solutions of best linear unbiased prediction (BLUP) using microbiome relationship matrix.

Results showed substantial contribution of the microbial variance (microbiability) on the feed efficiency related traits, and negligible effects on other performance traits, especially when the additive genetic effect was included in the linear mixed models. The microbiability estimates were lower than heritability values for all traits. Bivariate analyses showed a high microbial correlation between the feed efficiency traits. The MWAS using single-OTU regression method and back solving of BLUP solutions had high consistency, however, the detection powers were lower with the joint MWAS estimations resulting from back solving of the BLUP solutions. Poor values of the microbiability for performance traits did not seem to affect the detection power. The OTUs associated with the studied traits were annotated to the *Lachnospiraceae*, *Ruminococcaceae*, *Prevotellaceae* and *Streptococcaceae* families that are mainly involved in producing short-chain fatty acids and digestive enzymes. These detected taxonomic levels can be considered as future biomarkers in the improvement programs of feed efficiency of pigs.

This chapter is presented as a journal paper to be submitted to the GSE journal. In addition, the content of the chapter has been submitted for an oral presentation to the EAAP-2021 (Appendix 4.1).

4.2. Article III: Microbiability and microbiome wide association studies for feed efficiency and performance traits in pigs

Amir Aliakbari[1*], Olivier Zemb[1], Céline Barilly[1], Yvon Billon[2], Hélène Gilbert[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France

[2]GenESI, INRAE, 17700 Surgères, France

*Corresponding author: amir.aliakbari@inrae.fr

4.2.1. Abstract

Recent advances in obtaining microbiota information enable surveying the interplay between complex traits and the microbial community of the gastrointestinal tract (GIT). The objective of the present study was to investigate the contribution of faecal microbial variants to feed efficiency and other performance traits including average daily gain (ADG), back fat thickness (BFT), daily feed intake (DFI), feed conversion ratio (FCR), and residual feed intake (RFI) using data from two experimental pig lines that were divergent for feed efficiency. Microbiome wide association analyses (MWAS) were also run using two methods of single-OTU regression and back solving of solutions of best linear unbiased prediction using microbiome relationship matrix. The microbiabilities (m^2) obtained from linear animal models accounting for the genetic background of the hosts using the Bayesian approach. The h^2 posterior means were moderate for all traits and ranged from 0.31 ± 0.13 for FCR to 0.51 ± 0.10 for BFT. The m^2 posterior means of 0.11 ± 0.09 for RFI, and 0.20 ± 0.11 for FCR, 0.04 ± 0.03 for DFI, 0.03 ± 0.03 for ADG and 0.02 ± 0.03 for BFT were obtained. All traits showed lower m^2 than h^2 values and omitting the additive genetic effect resulted in higher residual variances. Bivariate analyses showed a high microbial correlation between the feed efficiency traits (0.70 ± 0.34). The two approaches used for MWAS showed similar results. However, significance levels of OTUs estimates were slightly different between the two methods, and the single-regression method showed higher significance. For RFI, the single-OTU regression showed three suggestive OTUs, whereas the back solving method showed one significant and one suggestive OTU. Both approaches showed one significant OTU for FCR and BFT. For DFI, the single-OTU regression showed two significant and one suggestive OTU, whereas with the back solving method one significant and one suggestive were found. Finally, for ADG, none of the methods pointed out significant or suggestive tests. The 8 OTUs with significant or suggestive effects on the five traits belonged to the *Streptococcaceae*, *Prevotellaceae*, *Ruminococcaceae*, and *Lachnospiraceae* families that are mainly involved in producing short-chain fatty acids and digestive enzymes. Our results showed substantial effects of the microbial variance on the feed efficiency related traits and negligible effects on performance traits. These results are confirmed the association between microbial community and complex phenotypes and detected

taxonomies can be considered as future biomarkers in the improvement programs of feed efficiency of pigs.

4.2.2. Introduction

Recent advances in obtaining microbiota information enable surveying the interplay between complex traits and the microbial community of the gastrointestinal tract (GIT) in animals and humans. This is especially essential in the pig industry where previous studies wildly revealed substantial contribution of the gut microbiome to the variability of feed efficiency in pigs (Bergamaschi et al., 2020; Camarinha-Silva et al., 2017). From the quantitative genetics perspective, the effect of the microbiome on a trait can be quantified by the microbiability, which is the proportion of phenotypic variance of the traits explained by the entire microbial community. Estimating the microbiability requires a microbial relationship matrix between different host animals (Difford et al., 2016). With such approach, Camarinha-Silva et al. (2017) reported higher microbiability for feed conversion ratio (FCR) (0.21 ± 0.14) and feed intake (0.16 ± 0.10) than the heritability of these traits. Similarly, a recent study revealed variation in the contribution of the microbiome to the meat quality and carcass composition traits in crossbred pigs over time, with increased contribution of microbiota to trait variability from weaning to off-test for the majority of the traits (Khanal et al., 2019), and higher microbiability than heritability for some traits, particularly at the off-test stage. In contrast, Tang et al. (2020) obtained a lower microbiability than the heritability for body weight (BW), average daily gain (ADG), backfat thickness (BFT), and intramuscular fatness using samples taken from five different points of the gut. Overall, these studies highlighted the importance and the high impact of variation derived from gut microbiome composition on the variation of different performance traits. Similar to the genome-wide association studies, microbiome variants can be considered as potential markers of the desired complex traits, and their associations can be identified through the microbiome-wide association studies (MWAS) (Difford et al., 2018). In an early MWAS investigation in the Piétrain pig breed, few outliers of marginal OTU effects were detected for ADG, FCR, and feed intake, and the authors concluded that these traits could have a polymicrobial nature (Camarinha-Silva et al., 2017). To the best of our knowledge, except the MWAS conducted by Camarinha-Silva

et al. (2017) there is no other published literature on this topic in swine, despite numerous examples in human, and few in other livestock species (Difford et al., 2018; Vollmar et al., 2020). The main objective of the present study was to investigate the contribution of faecal microbial variants to feed efficiency and other performance traits including ADG, BFT, daily feed intake (DFI), FCR, and residual feed intake (RFI) using data from two experimental pig lines that were divergent for feed efficiency. Before running association analyses, microbiabilities of traits were obtained using animal models that accounted for the genetic background of the hosts.

4.2.3. Materials and Methods

4.2.3.1. Population structure, studied traits and sampling

Phenotypic records were collected from two experimental French Large White pig lines. The lines were developed over 10 generations of divergent selection for RFI during 18 years at INRAE (UE GenESI, Surgères, France, 1999 to 2017, https://doi.org/10.15454/1.5572415481185847E12). The structure of the data and selection process of the lines has been described in Gilbert et al. (2017) and Aliakbari et al. (2020). Artificial insemination was used to obtain the G0 individuals from 60 F0 sows and boars. In G0, among the 116 candidates tested for RFI, 6 extreme low RFI (LRFI) and 6 extreme high RFI (HRFI) boars were selected as founder animals of each line. Random matings were implemented between the selected animals and 70 G0 gilts (equally distributed between the two lines) to produce generation G1. The same procedure with 96 tested boars per line was repeated to produce G1 to G10. Selection candidates had records for feed intake and feed efficiency traits, growth rate, and live body composition traits from 35 to 95 kg of body weight (BW). Additional females and castrated males had records to evaluate correlated responses to selection for growth rate, feed efficiency and carcass composition traits at each generation (response animals), with records from 10 weeks of age until slaughter (105 kg BW until G5 and 115 kg BW afterward). In all generations boars were selected based on a phenotypic index combining daily feed intake (DFI) and average daily gain (ADG) between 35 and 95 kg BW, and backfat thickness (BFT) at 95 kg BW (Gilbert et al., 2007), as DFI (g/d) − (1.06 × ADG (g/d)) − (37 × BFT (mm)). There was no selection on the sows, which were distributed in

two herds of birth with equal numbers of LRFI and HRFI sows in the two herds. After weaning (28 days of age), all pigs were penned in the same herd, in groups of 24, per line and sex. At 10 weeks of age, pigs from each pen were distributed in two growing-finishing pens (n=12 per pen). There were four pens per contemporary group (CG) and at least eight CG tested per generation over both lines (4 CG of candidates to selection and 4 CG of response animals). Growing-finishing pens were equipped with single-place electronic feeders ACEMA64 (ACEMO, France) to record individual feed intake. A pelleted diet based on cereals and soya bean meal was available ad libitum, and contained 10 MJ net energy (NE)/kg and 160 g CP/kg, with a minimum of 0.80 g digestible Lys/MJ NE. Animals had free access to water at all stages. Complete pedigree information was registered, starting at least one generation before F0 ancestors until G10.

Two different multiple linear regression equations, considering the test differences in candidates to selection and response animals, were used to compute realized RFI (Gilbert et al., 2007). The RFI for selection candidates was defined as the residuals of the regression of DFI on ADG and BFT (measured by ultrasounds). For response animals, the RFI equation included AMBW, ADG, carcass BFT (carcBFT) and lean meat content (LMC; computed from cut weights). In both models fixed effects of pen size and CG were fitted, and the fixed effect of sex and covariate of BW at the beginning of the test were added for response animals. Feed conversion ratio (FCR) was computed based on the corresponding test period of the two groups of animals. In this study, standardized phenotypes of RFI, FCR, DFI, ADG, and BFT were computed for both selection candidates and response animals and were analysed, as previously proposed in Aliakbari et al. (2020).

Faecal samples of 604 animals from G9 and G10 of the LRFI and HRFI lines were collected to obtain the gut microbial information. The samples collected in G9 were from selection candidates (boars) and the samples collected in G10 were from response animals (females and castrated males). Samplings were done at 15 weeks of age. Immediately after collection, the samples were homogenized and placed in dry ice, before storage at -80° C until DNA extraction (see next section). The descriptive information of the five traits from these individuals are given in Table 4-1.

Table 4-1. Descriptive statistics of data structure for the studied traits

Trait	Number	Min	Max	Average	SD
RFI	522	-0.38	0.39	0.00	0.15
FCR	548	1.604	3.928	2.779	0.333
DFI	542	1.37	2.95	2.20	0.29
ADG	575	0.514	1.011	0.776	0.079
BFT	541	9.82	46.56	23.28	10.02

ADG average daily gain (kg/day), *BFT* backfat thickness (mm), *DFI* daily feed intake (kg/day), *FCR* feed conversion ratio (kg/kg), *RFI* residual feed intake (kg/day)

4.2.3.2. Microbial information

The Quick-DNA™ Faecal Microbe Miniprep Kit™ (Zymo Research, Freiburg, Germany) was used to extract microbial DNA based on a 15 min bead-beating step at 30 hertz. Amplification of the V3-V4 region of the 16S rRNA gene obtained from diluted genomic DNA was done using two primers of F343 (CTTTCCCTACACGACGCTCTTCCGATCTTACGGRAGGCAGCAG) and R784 (GGAGTTCAGACGTGTGCTCTTCCGATCTTACCAGGGTATCTAATCCT) in 30 cycles and annealing temperature of 65 °C. To assemble pair-end sequences the Flash software v1.2.6 (Magoc & Salzberg, 2011) was used with at least 10-bp overlap between the forward and reverse sequences and allowing 10% mismatch. Single multiplexing was performed using an in-house 6 bp index, which was added to R784 during a second PCR with 12 cycles using forward primer (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC) and reverse primer (CAAGCAGAAGACGGCATACGAGAT-index-GTGACTGGAGTTCAGACGTGT). The resulted PCR products were then purified and loaded to the Illumina MiSeq cartridge based on the instructions of manufacturer. Quality of runs were internally checked using PhiX, and each pair-end sequence was assigned to its sample using the integrated index, with the bcl2fastq Illumina software. The sequences were submitted to the Short-Read Archive with accession number SRP124929. Filtering and trimming of sequences of high quality was applied to the reads with the DADA2 package in the R software (Callahan et al., 2016) with the following parameters: maxN=0, maxEE=2, truncQ=2, trimleft=17. Chimera were removed with the consensus method to obtain the final OTU abundance

table. No further clustering was applied, therefore operational taxonomic units (OTUs) were equivalent to amplicon sequence variants in this study. This step was followed by taxonomic annotation using the assignTaxonomy function of dada2 with the Silva Dataset v132 (Quast et al., 2013).

After rarefication of the abundance table to 9000 counts per sample and discarding 16 samples that contained fewer reads than the indicated counts, the final table contained 5689 OTUs for 588 samples (295 LRFI and 293 HRFI pigs). Finally, following Rothschild et al. (2018), OTUs in the rarefied table were filtered for more than 1% non-zero values across sampled animals, which diminished the number of OTUs to 2630.

4.2.3.3. Statistical analyses

4.2.3.3.1. Estimation of variance components

For all traits, four univariate linear models were applied to evaluate their goodness of fit regarding the microbiome effect. Therefore, the comparisons were between the models with and without the microbiome effect with degree of freedom equal to one, i.e. model 1 with 2, and model 3 with 4. The models were run in the Bayesian framework, so comparisons were based on the deviance information criterion (DIC; Spiegelhalter et al. (2002)) and estimations of variance components were obtained from Bayesian inference. The models were as in the following:

1) $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$
2) $\mathbf{y} = \mathbf{Xb} + \mathbf{Z_2 m} + \mathbf{e}$
3) $\mathbf{y} = \mathbf{Xb} + \mathbf{Z_1 a} + \mathbf{e}$
4) $\mathbf{y} = \mathbf{Xb} + \mathbf{Z_1 a} + \mathbf{Z_2 m} + \mathbf{e}$

where $\mathbf{y}$ is the vector of observations of the each of the five traits, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{a}$ is the vector of random additive genetic effects, $\mathbf{m}$ is the vector of random microbiome effects, and $\mathbf{e}$ is the vector of random residuals. $\mathbf{X}$, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the incidence matrices for $\mathbf{b}$, $\mathbf{a}$ and $\mathbf{m}$. The distributions assumed for the random terms are $\mathbf{a} \sim N(0, \mathbf{A}\sigma_a^2)$, $\mathbf{m} \sim N(0, \mathbf{M}\sigma_m^2)$ and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, and σ_a^2, σ_m^2 and σ_m^2 are the variances of direct additive genetic effect, microbiome effect and residual effect, respectively.

$\mathbf{I}$ denoted the identity matrix. The pedigree relationship matrix ($\mathbf{A}$) contained the 588 animals with microbiota data, plus 6705 ancestors (parents from generations G0 to G8 of the lines, plus their ancestors in common original population). $\mathbf{M}$ is the microbial relationship matrix (Camarinha-Silva et al., 2017) and is defined as $\mathbf{M} = \frac{\mathbf{Z_3 Z_3}'}{\mathrm{k}}$, where $\mathbf{Z_3}$ is a matrix with dimension of n x k, where n is the number of animals with microbiome information and k is the number of OTUs. Elements of the $\mathbf{Z_3}$ matrix are the standardized individual abundance of each OTU j for individual i, according to the following equation:

$$z_{3\,ij} = \frac{\log(P_{ij}) - \overline{\log(P_j)}}{sd(\log(P_j))} \qquad \textbf{(1)}$$

Where P_{ij} is the abundance of OTU j for individual i, and P_j is the vector of abundances of the j[th] OTU.

The fixed environmental factors fitted in the model were the pen size (5 levels), herd of birth (two levels), sex (three levels), and contemporary groups (CG, 109 levels). Their significance ($p < 0.05$) on the five traits was tested in preliminary linear models.

In addition, in order to assess the microbial correlations (r_m) between the traits with consideration of additive genetic effect, bivariate analyses with model 4 were run. In this case, the distributions assumed for the random terms were $\mathbf{a} \sim N(0, \mathbf{G_0} \otimes \mathbf{A})$, $\mathbf{m} \sim N(0, \mathbf{R_m} \otimes \mathbf{I})$ and $\mathbf{e} \sim N(0, \mathbf{R_e} \otimes \mathbf{I})$, where $\mathbf{G_0} = \begin{bmatrix} \sigma^2_{a_i} & \sigma_{a_{ij}} \\ \sigma_{a_{ji}} & \sigma^2_{a_j} \end{bmatrix}$ is a 2×2 symmetric (co)variance matrix of direct additive genetic effects including the previously defined genetic variances and the genetic correlation $r_{g_{ij}} = \frac{\sigma_{a_{ij}}}{\sigma_{a_i}\sigma_{a_j}}$ between each pair of traits i and j, and similarly $\mathbf{R_m} = \begin{bmatrix} \sigma^2_{m_i} & \sigma_{m_{ij}} \\ \sigma_{m_{ji}} & \sigma^2_{m_j} \end{bmatrix}$, with $r_{\mathbf{m_{ij}}} = \frac{\sigma_{m_{ij}}}{\sigma_{m_i}\sigma_{m_j}}$, and $\mathbf{R_e} = \begin{bmatrix} \sigma^2_{e_i} & \sigma_{e_{ij}} \\ \sigma_{e_{ji}} & \sigma^2_{e_j} \end{bmatrix}$ are 2×2 symmetric (co)variances matrices of microbiome and residual effects, respectively.

The analyses were performed using the GIBBSF90 software (Misztal et al., 2018). In total, 100,000 samples were generated to obtain the posterior distributions of the parameters of the model, and a burn-in period of 15,000 samples and thinning interval of

10 were considered. The convergence was verified through visual inspection of trace sample plots.

4.2.3.3.2. Microbiome wide association studies

The objective in this step was to identify OTUs that have significant associations with the studied traits. Two separate approaches were used.

a) Using single-OTU regressions: first, single-OTU regression analyses were applied to test the effect of the 2630 OTUs one at a time and obtain associated p-value, which is the most common approach (Difford et al., 2018). The model defined for these analyses was same as model (3) except that OTUs were fitted as fixed covariates in addition to the other fixed effects. The AIREMLF90 software (Misztal et al., 2018) was used to run the BLUP method. The p-values of resulted regression coefficients were obtained by converting coefficient estimates and their standard error into corresponding Z-scores and applying a chi² test.

b) Using back solving of BLUP solutions: in an alternative approach, contributions of microbiota to the variance of each OTU were retrieved from the microbiability model similar to what explained by Stranden and Garrick (2009) and Gualdron Duarte et al. (2014) to obtain solutions and prediction error variances for SNP markers from genomic-BLUP solutions. Such back solving is often used in the SNP GWAS literature, but appeared only recently for microbiota analyses (Vollmar et al., 2020).

Solutions for OTUs ($\widehat{\mathbf{OTU}}$) can be obtained if the assumptions of $\sigma^2_{\mathrm{OTU}} = \sigma^2_{\mathrm{m}}/\mathrm{k}$ and $\mathbf{D} = \mathbf{I}\sigma^2_{\mathrm{m}}/\mathrm{k}$ hold and thus:

$$\mathbf{Z_3 D Z_3}' = \mathbf{M}\sigma^2_{\mathrm{m}} \qquad \textbf{(2)}$$

Therefore, solutions for OTUs effects given the microbiome solutions can be achieved as in the following (Stranden & Garrick, 2009):

$$\mathbf{E(OTU|y)} = \widehat{\mathbf{OTU}}|\widehat{\mathbf{m}} = \mathbf{D Z_3}'(\mathbf{Z_3 D Z_3}')^{-1}\widehat{\mathbf{m}} = \tfrac{1}{\mathrm{k}}\mathbf{Z_3}'\mathbf{M}^{-1}\widehat{\mathbf{m}} \qquad \textbf{(3)}$$

And the variance of OTUs solutions is defined as (Gualdron Duarte et al., 2014):

$$\mathbf{var}\left(\widehat{\mathbf{OTU}}\right) = \mathbf{var}\left(\tfrac{1}{\mathbf{k}}\mathbf{Z_3}'\mathbf{M}^{-1}\widehat{\mathbf{m}}\right) = \tfrac{1}{\mathbf{k}}\mathbf{Z_3}'\mathbf{M}^{-1}\mathbf{var}(\widehat{\mathbf{m}})\,\mathbf{M}^{-1\prime}\mathbf{Z_3}\tfrac{1}{\mathbf{k}} \qquad \textbf{(4)}$$

The predictor error variance (PEV) of $\widehat{\mathbf{m}}$ is equal to:

$$\mathbf{PEV}(\widehat{\mathbf{m}}) = \mathbf{var}(\mathbf{m} - \widehat{\mathbf{m}}) = \mathbf{var}(\mathbf{m}) - \mathbf{var}(\widehat{\mathbf{m}}) = \mathbf{C}^{\mathbf{mm}}\boldsymbol{\sigma}_{\mathbf{e}}^{\mathbf{2}} \quad \textbf{(5)}$$

Therefore

$$\mathbf{var}(\widehat{\mathbf{m}}) = \mathbf{var}(\mathbf{m}) - \mathbf{C}^{\mathbf{mm}}\boldsymbol{\sigma}_{\mathbf{e}}^{\mathbf{2}} = \mathbf{M}\boldsymbol{\sigma}_{\mathbf{m}}^{\mathbf{2}} - \mathbf{C}^{\mathbf{mm}}\boldsymbol{\sigma}_{\mathbf{e}}^{\mathbf{2}} \quad \textbf{(6)}$$

where $\mathbf{C}^{\mathbf{mm}}$ are the diagonal elements of the sub-matrix corresponding to the microbiome random effect from the inverse of coefficient matrix of the mixed model equations.

Finally:

$$\mathbf{var}\left(\widehat{\mathbf{OTU}}\right) = \frac{\mathbf{1}}{\mathbf{k}}\mathbf{Z_3}'\,\mathbf{M}^{-\mathbf{1}}(\mathbf{M}\boldsymbol{\sigma}_{\mathbf{m}}^{\mathbf{2}} - \mathbf{C}^{\mathbf{mm}}\boldsymbol{\sigma}_{\mathbf{e}}^{\mathbf{2}})\,\mathbf{M}^{-\mathbf{1}}\mathbf{Z_3}\,\frac{\mathbf{1}}{\mathbf{k}} \quad \textbf{(7)}$$

Then, the Z-score for each OTUs solution j can be obtained as:

$$\mathrm{Z\_score}_j = \frac{\widehat{\mathrm{OTU}}_{\mathrm{j}}}{\sqrt{\mathrm{var}\left(\widehat{\mathrm{OTU}}\right)_{\mathrm{j}}}} \quad \textbf{(8)}$$

The corresponding p-values can then be calculated by applying a Chi² test to these Z-scores.

The back solving method was run using a local script for construction and solving of the mixed model equations based on the variance component estimates of the model 4 for each trait.

4.2.3.3.3. Significance threshold for MWAS

To estimate the number of independent tests and calculate the significance thresholds for the MWAS, a principal component analysis (PCA) was applied to the correlation matrix of OTUs ($\mathbf{Z_3}'\mathbf{Z_3}_{\mathbf{2630}\times\mathbf{2630}}$) to control the family-wise type I error rate of 5% while accounting for multiple testing, as proposed by Gao et al. (2008). The PCA showed that 428 eigenvalues captured 99.5% of the variability in the correlation matrix. Based on this, to test the significance of OTUs effects, two cut-off points for a significance 5% threshold (–log10(0.05/428)) and for a suggestive 10% threshold –log10(0.10/428) were utilized.

4.2.4. Results

4.2.4.1. Estimation of variance components

The results of univariate analyses of the five studied traits with models including the microbiome effect are presented in Table 4-2, as posterior means ± posterior 95% confidence intervals of each variance component. The comparisons of the DIC values showed a consistent improvement of models from model 1 to model 4, which had the smallest DIC for all traits. The h^2 posterior means were moderate for all traits and ranged from 0.31 ± 0.13 for FCR to 0.51 ± 0.10 for BFT, with no difference between estimates from models 3 and 4. The microbiome variance obtained with models 2 and 4 showed substantial contribution to the phenotypic variance of feed efficiency related traits, with m^2 of 0.22 ± 0.11 to 0.20 ± 0.11 for FCR, and 0.12 ± 0.09 and 0.11 ± 0.09 for RFI, respectively. In contrast, phenotypic variances of DFI, BFT and ADG showed less influence of the microbiome variance, with posterior means lower than 0.06 ± 0.06 for DFI and ADG with the two models, and a change from 0.11 ± 0.06 (model 2) to 0.02 ± 0.03 (model 4) for BFT, i.e. mainly not differing from zero. All traits showed m^2 posterior means lower than h^2 posterior means, and omitting the additive genetic effect in the models 1 and 2 resulted in higher residual variances in comparison to models 3 and 4 for all traits.

The results of bivariate analyses between the traits with model 4 are given in Table 4-3. The h^2 and m^2 estimates of the traits in these analyses were similar to estimates obtained from univariate analyses. The r_m estimates between the traits ranged from -0.37 ± 0.56 for DFI and ADG to 0.96 ± 0.11 for ADG and BFT. Except the r_m estimate between RFI and FCR (0.70 ± 0.34), other estimates, given the low microbiability estimates of the traits, were estimated with very low reliabilities.

Table 4-2. Posterior means (± posterior standard deviation) of variance components, heritability and microbiability values of production traits using the four models, and corresponding deviance information criterion (DIC) of each model

Trait	Model	σ^2_g	σ^2_m	σ^2_e	σ^2_p	h^2	m^2	DIC
RFI	(1)	-	-	0.020 ± 0.001	0.020 ± 0.001	-	-	-343032380251
	(2)	-	0.002 ± 0.002	0.017 ± 0.002	0.020 ± 0.001	-	0.12 ± 0.09	-376715009798
	(3)	0.006 ± 0.002	-	0.014 ± 0.002	0.020 ± 0.001	0.32 ± 0.10	-	-481592517646
	(4)	0.006 ± 0.002	0.002 ± 0.002	0.012 ± 0.002	0.020 ± 0.001	0.30 ± 0.10	0.11 ± 0.09	-540693245186
FCR	(1)	-	-	0.062 ± 0.004	0.062 ± 0.004	-	-	-64257520104
	(2)	-	0.014 ± 0.008	0.051 ± 0.007	0.065 ± 0.005	-	0.22 ± 0.11	-78170373902
	(3)	0.024 ± 0.010	-	0.043 ± 0.007	0.067 ± 0.005	0.35 ± 0.13	-	-93388582161
	(4)	0.022 ± 0.010	0.014 ± 0.008	0.032 ± 0.009	0.070 ± 0.006	0.31 ± 0.13	0.20 ± 0.11	-122763803675
DFI	(1)	-	-	0.051 ± 0.003	0.051 ± 0.003	-	-	-89420640342
	(2)	-	0.003 ± 0.003	0.050 ± 0.004	0.052 ± 0.003	-	0.06 ± 0.06	-93351683359
	(3)	0.030 ± 0.010	-	0.030 ± 0.006	0.056 ± 0.005	0.50 ± 0.13	-	-167531111339
	(4)	0.030 ± 0.010	0.002 ± 0.002	0.030 ± 0.007	0.060 ± 0.005	0.48 ± 0.14	0.04 ± 0.03	-173089799612
ADG	(1)	-	-	0.0051 ± 0.0003	0.0051 ± 0.0003	-	-	-255536305168
	(2)	-	0.0002 ± 0.0003	0.0050 ± 0.0004	0.0051 ± 0.0003	-	0.05 ± 0.05	-265759148148
	(3)	0.0024 ± 0.0009	-	0.0030 ± 0.0006	0.0054 ± 0.0005	0.45 ± 0.13	-	-440920106295
	(4)	0.0030 ± 0.0008	0.0001 ± 0.0002	0.0030 ± 0.0006	0.0055 ± 0.0005	0.47 ± 0.12	0.03 ± 0.03	-471061077561
BFT	(1)	-	-	8.854 ± 0.570	8.854 ± 0.570	-	-	-525486682
	(2)	-	0.100 ± 0.610	8.057 ± 0.680	9.055 ± 0.604	-	0.11 ± 0.06	-575243859
	(3)	4.754 ± 1.301	-	4.695 ± 0.872	9.450 ± 0.750	0.50 ± 0.11	-	-999015096
	(4)	4.980 ± 1.280	0.228 ± 0.314	4.424 ± 0.878	9.636 ± 0.760	0.51 ± 0.10	0.02 ± 0.03	-1071710845

σ^2_g: genetic variance, σ^2_m: microbiome variance, σ^2_e: residual variance, σ^2_p: phenotypic variance, h^2: heritability, m^2: microbiability

Table 4-3. Posterior means (± posterior standard deviation) of the main parameters obtained from bivariate analyses between traits with model 4

Trait 1	Trait 2	h^2_{T1}	h^2_{T2}	m^2_{T1}	m^2_{T2}	r_{g12}	r_{m12}
RFI	FCR	0.35 ± 0.11	0.38 ± 0.13	0.16 ± 0.10	0.23 ± 0.10	0.66 ± 0.16	0.70 ± 0.34
	DFI	0.29 ± 0.09	0.54 ± 0.13	0.18 ± 0.12	0.06 ± 0.04	0.63 ± 0.17	0.71 ± 0.47
	ADG	0.33 ± 0.10	0.51 ± 0.13	0.09 ± 0.08	0.10 ± 0.05	0.00 ± 0.27	-0.54 ± 0.60
	BFT	0.29 ± 0.10	0.52 ± 0.11	0.17 ± 0.08	0.05 ± 0.03	0.02 ± 0.27	-1.00 ± NE
FCR	DFI	0.32 ± 0.12	0.52 ± 0.12	0.28 ± 0.10	0.08 ± 0.03	0.44 ± 0.25	0.99 ± NE
	ADG	0.38 ± 0.13	0.51 ± 0.13	0.22 ± 0.10	0.11 ± 0.05	-0.25 ± 0.21	-0.91 ± 0.18
	BFT	0.34 ± 0.11	0.49 ± 0.10	0.23 ± 0.09	0.05 ± 0.04	0.52 ± 0.20	0.40 ± 0.64
DFI	ADG	0.48 ± 0.13	0.49 ± 0.12	0.04 ± 0.04	0.12 ± 0.06	0.62 ± 0.16	-0.37 ± 0.56
	BFT	0.50 ± 0.13	0.51 ± 0.10	0.04 ± 0.04	0.06 ± 0.04	0.61 ± 0.15	0.50 ± 0.58
ADG	BFT	0.49 ± 0.14	0.50 ± 0.11	0.04 ± 0.05	0.06 ± 0.05	0.30 ± 0.20	0.96 ± 0.11

h^2_{T1}: heritability of first trait, h^2_{T2}: heritability of second trait, m^2_{T1}: microbiability of first trait, m^2_{T2}: microbiability of second trait, r_{g12}: genetic correlation, r_{m12}: microbial correlation, NE: not estimable

4.2.4.2. Microbiome wide association studies (MWAS)

The two approaches used for MWAS, i.e. single OTU regression and back solving of BLUP solutions, showed similar results. However, significance levels of OTUs estimates were slightly different between the two methods where the single-regression method showed higher significance. Results of MWAS with single OTU regression are shown in Figure 4-1 and Figure 4-2, and those from the back solving approach are given in Figure S1. There was no common significant or suggestive OTU between the traits. For RFI, the single-OTU regression showed three suggestive OTUs (OTU391, OTU1749 and OTU2280), whereas the back solving method showed one significant (OTU391), and one suggestive OTU (OTU1749). Both approaches showed one significant OTU for FCR (OTU1768) and BFT (OTU2934). For DFI, the single-OTU regression showed two significant (OTU694, OTU1619) and one suggestive OTU (OTU2678), whereas with the back solving method one significant (OTU694) and one suggestive (OTU1619) were found. Finally, for ADG, none of the methods pointed out significant or suggestive tests.

The 8 OTUs with significant or suggestive effects on the five traits belonged to the *Streptococcaceae* (1 OTU), *Prevotellaceae* (3), *Ruminococcaceae* (3) and *Lachnospiraceae* (1) families (Table 4-4). From these, the 6 OTUs with identified genus belonged to different genera. All these genera had more than 85% of zeros.

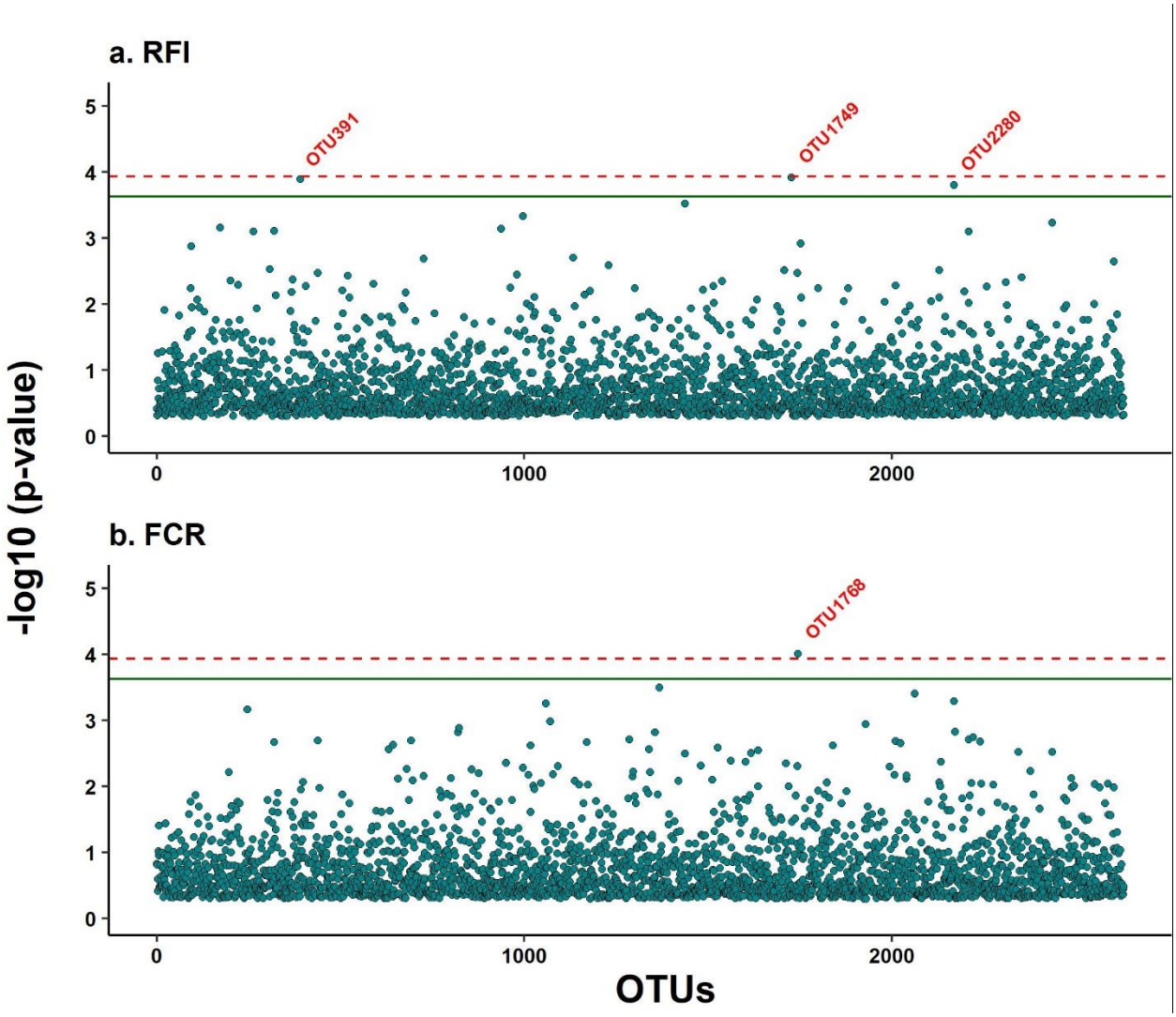


Figure 4-1. Results of microbiome wide association between operational taxonomic units and residual feed intake (a) and feed conversion ratio (b)

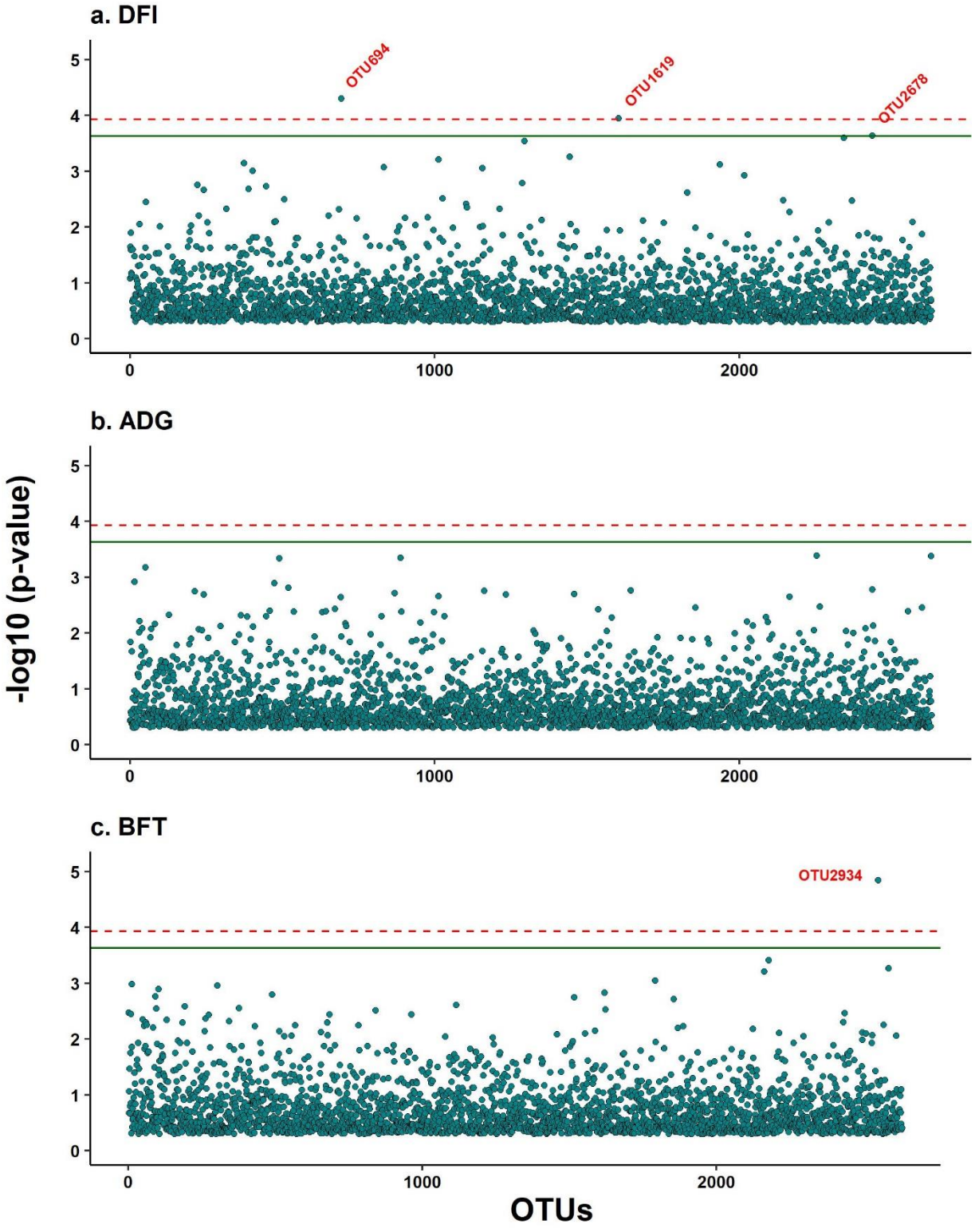


Figure 4-2. Results of microbiome wide association study between operational taxonomic units and daily feed intake (a), average daily gain (b) and back fat thickness (c)

Table 4-4. Taxonomy and descriptive statistics (minimum (Min), maximum (Max), mean and standard deviation (*SD*)) of the OTUs with significant/suggestive associations with the five studied traits

Trait	OTU	Kingdom	Phylum	Class	Order	Family	Genus	%Zero	Min	Max	Average	SD
RFI	**OTU391**	*Bacteria*	*Firmicutes*	*Bacilli*	*Lactobacillales*	*Streptococcaceae*	*Streptococcus*	92.69	0	113	2.33	11.37
	OTU1749	*Bacteria*	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Prevotellaceae*	*Prevotella_9*	98.81	0	32	0.11	1.46
	OTU2280	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Ruminococcaceae*	*unknown*	97.79	0	9	0.06	0.53
FCR	**OTU1768**	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Ruminococcaceae*	*Ruminococcus_1*	98.47	0	62	0.14	2.59
DFI	**OTU694**	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Ruminococcaceae*	*Ruminococcaceae_UCG-014*	85.03	0	65	1.00	4.32
	OTU1619	*Bacteria*	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Prevotellaceae*	*Alloprevotella*	94.90	0	17	0.19	1.17
	OTU2678	*Bacteria*	*Firmicutes*	*Clostridia*	*Clostridiales*	*Lachnospiraceae*	*XBB1006*	98.30	0	4	0.03	0.28
BFT	**OTU2934**	*Bacteria*	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Prevotellaceae*	*unknown*	98.64	0	3	0.02	0.19

4.2.5. Discussion

4.2.5.1. Estimation of variance components

Previous studies in pigs revealed that part of the microbial community are heritable (Camarinha-Silva et al., 2017; Difford et al., 2018; Mach et al., 2015), which would provide a stability of microbial components presence in gastrointestinal tract (GIT) across generations. Such stability of the microbial community could favor their contribution to the variability of the phenotypes of the host animals. Therefore, in the present study, gut microbial information of two divergent pig lines was fitted into the linear animal mixed models to predict its contribution to the phenotypic variance of feed efficiency and other performance traits. The analyses showed substantial effects of the microbial variance on the feed efficiency related traits. The m^2 obtained for RFI in our study was lower than the reported value (0.45 ± 0.15) by Weishaar et al. (2020). For FCR, m^2 values were in the range of the reports of 0.21 ± 0.14 from Camarinha-Silva et al. (2017) and 0.13 ± 0.10 from Weishaar et al. (2020).

For other performance traits, low estimates of m^2 were obtained, despite the lower DIC obtained with models 2 and 4 compared to models 1 and 3, respectively. In the study conducted by Camarinha-Silva et al. (2017) a non-significant m^2 of 0.16 ± 0.10 for feed intake was reported. This estimate was higher than our estimated m^2 with models 2 and 4 for DFI, but the confidence intervals would overlap. Camarinha-Silva et al. (2017) and Weishaar et al. (2020) reported moderate m^2 of 0.28 ± 0.13 and 0.24 ± 0.11, respectively, for daily gain, which were higher than our posteriori mean values for ADG. Khanal et al. (2019) in a study on the microbiability of meat quality and carcass composition traits in swine found an increasing m^2 of back fat depth by increasing age at sampling and reported m^2 of 0.01 ± 0.02 at weaning, 0.12 + 0.04 at mid-test and 0.25 ± 0.04 at off-test. Our estimates of m^2 for BFT with the full model is comparable with their report at weaning, whereas the sampling time of our study would be equivalent to their mid-test sampling. The three previous studies had different genetic types (Piétrain or commercial crossbreds), sample sizes and time of collection that could explain part of the difference between the studies. In addition, differences in the bioinformatics processing of the sequences to obtain OTU tables remain a factor of heterogeneity between studies.

To our knowledge, except one study on meat quality and carcass composition traits (Khanal et al., 2019), there is no study in pigs reporting microbial correlations between the studied traits. The positive high estimated r_m between the RFI and FCR suggests that a common microbial community have influence on both traits. Khanal et al. (2019) observed a decrease in genomic correlations between traits with higher microbial correlation and they argued that genomic correlations among traits are partially due to the correlations among the gut microbiota composition. However, given that we have already observed significant genetic correlations between the microbial components and the studied traits (Aliakbari et al., 2021), a reverse hypothesis can be also relevant, such that part of the r_m between the traits could be due to the high genetic correlations between the traits. The change of genetic correlation from model 3 to model 4 was not available yet from our analysis, and a more complete dataset and a higher sequencing depth for the genetic and microbial analyses will be needed to clarify this issue.

4.2.5.2. Microbiome wide association results

As could be expected, the results of the two approaches used to detect the association between OTUs and the phenotypic traits had high consistency, with different powers. As a result, the single-OTU regression showed two more suggestive OTUs than the back solving method (8 versus 6). This difference could be due to the properties of the BLUP method, which tends to shrink the effect solutions toward the mean of the population. This shrinkage can potentially be passed to the OTU effect estimates after the back solving. Therefore, single marker regression is more powerful than BLUP based methods for association studies, as was already shown for SNP analyses. However, the number of computations in this approach is equal to the number of OTU being tested, which could be limiting for a vast number of OTU.

The consistency between the MWA results of the single-OTU regression and back solving approaches can be considered as a confirmation of the estimated microbiome variance with the full model. For example, the single-OTU regression approach did not point out different associations as compared to the back solving method using the low values of the microbiome variance from model 4 for DFI, ADG and BFT.

The significant tests suggest that the some OTUs and phenotypic traits are associated. These associations, in fact, indicates that phenotype observations differ among those OTUs. Even though we did not conduct separate analyses for each line, as they would have a limited power, the differing abundance of the significant OTUs together with phenotype could be a result of divergent selection, as some of these components were shown previously to differ at the genera level (Aliakbari et al., 2021). Further investigations would be needed to prove this assumption.

In our previous study at the genera level, using the same microbiome dataset, we showed significant genetic correlations between genera from the *Lachnospiraceae*, *Ruminococcaceae*, *Prevotellaceae* and *Streptococcaceae* families with RFI, DFI, and BFT (Aliakbari et al., 2021). Therefore, finding significant OTUs associated to the phenotypic traits from these families is consistent with our previous results and probably part of their association contributes to the genetic correlation of these families with the studied traits. Weishaar et al. (2020) also reported OTUs from *Lachnospiraceae* and *Prevotellaceae* families that showed strong effect on FCR and RFI. The *Prevotellaceae*, *Lachnospiraceae* and *Ruminococcaceae* families are involved in the digestion of fibrous material of the nutrients and finally provide short-chain fatty acids for the host (Biddle et al., 2013; Gardiner et al., 2020). Bacteria from the *Streptococcaceae* family are known as lactic acid producer bacteria (du Toit et al., 2014) that has an important role in the production of dietary enzymes, such as amylase, lipase, phytase, and protease (Kim et al., 2007). Therefore, the identified OTUs could have meaningful biological links for feed efficiency and other performance traits. If confirmed in more diverse conditions, these OTUs could be used as potential biomarkers in selection programs to improve the feed efficiency of pigs. In addition, how the use of microbiability in linear mixed models improve the prediction accuracies for selection remains to be assess for these traits. Finally, the genetic background of the identified OTUs should be studied to indicate the magnitude of their genetic control, so that distinction between microbial and genetic effects could be achieved.

4.2.6. Conclusion

Our results showed substantial effects of the microbial variance on the feed efficiency related traits and negligible effects on performance traits, especially when the genetic effects were included in the models. The microbiability values were lower than heritability values for all traits. A high microbial correlation between the feed efficiency traits was observed. Our results also showed that MWAS using single-OTU regression method and back solving of BLUP solutions have high consistency, but detection power was lower with the later approach. However, low values of microbiability did not seem to affect the detection power. The OTUs associated with the traits were annotated to the *Lachnospiraceae*, *Ruminococcaceae*, *Prevotellaceae* and *Streptococcaceae* families that are mainly involved in producing short-chain fatty acids and digestive enzymes. Finally, these results confirmed the existence of associations between microbial community and complex phenotypes, and the detected taxons could be considered as future biomarkers in improvement programs of feed efficiency of pigs.

4.2.7. References

Aliakbari, A., Delpuech, E., Labrune, Y., Riquet, J., & Gilbert, H. (2020). The impact of training on data from genetically-related lines on the accuracy of genomic predictions for feed efficiency traits in pigs. *Genetics Selection Evolution*, *52*(1), 57. https://doi.org/10.1186/s12711-020-00576-0

Aliakbari, A., Zemb, O., Billon, Y., Barilly, C., Ahn, I., Riquet, J., & Gilbert, H. (2021). Genetic relationships between feed efficiency and gut microbiome in pig lines selected for residual feed intake. *Journal of Animal Breeding and Genetics*.

Bergamaschi, M., Maltecca, C., Schillebeeckx, C., McNulty, N. P., Schwab, C., Shull, C., Fix, J., & Tiezzi, F. (2020). Heritability and genome-wide association of swine gut microbiome features with growth and fatness parameters. *Sci Rep*, *10*(1), 10134. https://doi.org/10.1038/s41598-020-66791-3

Biddle, A., Stewart, L., Blanchard, J., & Leschine, S. (2013). Untangling the genetic basis of fibrolytic specialization by Lachnospiraceae and Ruminococcaceae in diverse gut communities. *Diversity*, *5*(3), 627-640.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581-583. https://doi.org/10.1038/nmeth.3869

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S., & Bennewitz, J. (2017). Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. *Genetics*, *206*(3), 1637-1644. https://doi.org/10.1534/genetics.117.200782

Difford, G., Lassen, J., & Løvendahl, P. (2016). Genes and microbes, the next step in dairy cattle breeding. Book of Abstracts of the 67th Annual Meeting of the European Federation of Animal Science,

Difford, G. F., Plichta, D. R., Lovendahl, P., Lassen, J., Noel, S. J., Hojberg, O., Wright, A. G., Zhu, Z., Kristensen, L., Nielsen, H. B., Guldbrandtsen, B., & Sahana, G. (2018). Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genetics*, *14*(10), e1007580. https://doi.org/10.1371/journal.pgen.1007580

du Toit, M., Huch, M., Cho, G. S., & Franz, C. (2014). The genus Streptococcus. *Lactic Acid Bacteria*, 457-505. https://doi.org/doi.org/10.1002/9781118655252.ch28

Gao, X., Starmer, J., & Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology*, *32*(4), 361-369. https://doi.org/10.1002/gepi.20310

Gardiner, G. E., Metzler-Zebeli, B. U., & Lawlor, P. G. (2020). Impact of Intestinal Microbiota on Growth and Feed Efficiency in Pigs: A Review. *Microorganisms*, *8*(12), 1886. https://doi.org/10.3390/microorganisms8121886

Gilbert, H., Bidanel, J. P., Gruand, J., Caritez, J. C., Billon, Y., Guillouet, P., Lagant, H., Noblet, J., & Sellier, P. (2007). Genetic parameters for residual feed intake in growing pigs, with emphasis on genetic relationships with carcass and meat quality traits. *Journal of Animal Science*, *85*(12), 3182-3188. https://doi.org/10.2527/jas.2006-590

Gilbert, H., Billon, Y., Brossard, L., Faure, J., Gatellier, P., Gondret, F., Labussiere, E., Lebret, B., Lefaucheur, L., Le Floch, N., Louveau, I., Merlot, E., Meunier-Salaun, M. C., Montagne, L., Mormede, P., Renaudeau, D., Riquet, J., Rogel-Gaillard, C., van Milgen, J., Vincent, A., & Noblet, J. (2017). Review: divergent selection for residual feed intake in the growing pig. *Animal*, *11*(9), 1427-1439. https://doi.org/10.1017/S175173111600286X

Gualdron Duarte, J. L., Cantet, R. J., Bates, R. O., Ernst, C. W., Raney, N. E., & Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinformatics*, *15*(1), 246. https://doi.org/10.1186/1471-2105-15-246

Khanal, P., Maltecca, C., Schwab, C., Fix, J., & Tiezzi, F. (2019). Microbiability of meat quality and carcass composition traits in swine. *bioRxiv*, 833731.

Kim, E. Y., Kim, Y. H., Rhee, M. H., Song, J. C., Lee, K. W., Kim, K. S., Lee, S. P., Lee, I. S., & Park, S. C. (2007). Selection of Lactobacillus sp. PSC101 that produces active dietary enzymes such as amylase, lipase, phytase and protease in pigs. *Journal of General and Applied Microbiology*, *53*(2), 111-117. https://doi.org/10.2323/jgam.53.111

Mach, N., Berri, M., Estelle, J., Levenez, F., Lemonnier, G., Denis, C., Leplat, J. J., Chevaleyre, C., Billon, Y., Dore, J., Rogel-Gaillard, C., & Lepage, P. (2015). Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ Microbiol Rep*, *7*(3), 554-569. https://doi.org/10.1111/1758-2229.12285

Magoc, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957-2963. https://doi.org/10.1093/bioinformatics/btr507

[Record #60 is using a reference type undefined in this output style.]

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), 590-596. https://doi.org/10.1093/nar/gks1219

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., Avnit-Sagi, T., Lotan-Pompan, M., Weinberger, A., Halpern, Z., Carmi, S., Fu, J., Wijmenga, C., Zhernakova, A., Elinav, E., & Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, *555*(7695), 210-215. https://doi.org/10.1038/nature25973

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *64*(4), 583-616. https://doi.org/Doi 10.1111/1467-9868.00353

Stranden, I., & Garrick, D. J. (2009). Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *Journal of Dairy Science*, *92*(6), 2971-2975. https://doi.org/10.3168/jds.2008-1929

Tang, S., Xin, Y., Ma, Y., Xu, X., Zhao, S., & Cao, J. (2020). Screening of Microbes Associated With Swine Growth and Fat Deposition Traits Across the Intestinal Tract. *Front Microbiol*, *11*, 586776. https://doi.org/10.3389/fmicb.2020.586776

Vollmar, S., Wellmann, R., Borda-Molina, D., Rodehutscord, M., Camarinha-Silva, A., & Bennewitz, J. (2020). The Gut Microbial Architecture of Efficiency Traits in the

Domestic Poultry Model Species Japanese Quail (Coturnix japonica) Assessed by Mixed Linear Models. *G3 (Bethesda)*, *10*(7), 2553-2562. https://doi.org/10.1534/g3.120.401424

Weishaar, R., Wellmann, R., Camarinha-Silva, A., Rodehutscord, M., & Bennewitz, J. (2020). Selecting the hologenome to breed for an improved feed efficiency in pigs—A novel selection index. *Journal of Animal Breeding and Genetics*, *137*(1), 14-22.

4.2.8. Appendix 4.1

72. Annual meeting of the European Federation of Animal Science (EAAP), Virtual Meeting, Dec, 2021

Microbiability and microbiome-wide associations with feed efficiency and performance traits in pigs

Amir Aliakbari[1], Olivier Zemb[1], Céline Barilly[1], Yvon Billon[2], Hélène Gilbert[1]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France

[2]GenESI, INRAE, 17700 Surgères, France

The present study aimed at investigating in pigs the contribution of faecal microbial composition (microbiability) to feed efficiency and other performance traits including average daily gain (ADG), back fat thickness (BFT), daily feed intake (DFI), feed conversion ratio (FCR), and residual feed (RFI). The operational taxonomic units (OTU) abundances were obtained from 16S rRNA sequencing of fecal samples from about 550 pigs from two lines divergently selected for RFI. The microbiabilities (m^2) were obtained from mixed linear animal models accounting for the additive genetic background of the pigs using a Bayesian approach. Microbiome-wide association studies (MWAS) were run using single-OTU regressions or back solving the solutions of best linear unbiased predictions from the microbiome relationship matrix. The heritability posterior means (h^2) were moderate for all traits, ranging from 0.31 ± 0.13 for FCR to 0.51 ± 0.10 for BFT. The m^2 posterior means were 0.11 ± 0.09 for RFI, 0.20 ± 0.11 for FCR, 0.04 ± 0.03 for DFI, 0.03 ± 0.03 for ADG and 0.02 ± 0.03 for BFT. All traits showed lower m^2 than h^2 values. Omitting the additive genetic effect resulted in higher residual variances, and higher m^2 for BFT only (0.11 ± 0.06). The two approaches used for MWAS showed similar results, but the single-regression method had higher detection power. With this approach, three

suggestive OTUs were found for RFI, one significant OTU was found for FCR and BFT. For DFI two significant and one suggestive OTU were found. For ADG, no association was found. These 8 OTUs belonged to the *Streptococcaceae*, *Prevotellaceae*, *Ruminococcaceae*, and *Lachnospiraceae* families, mainly involved in producing short-chain fatty acids and digestive enzymes. Therefore, our results showed a substantial contribution of the microbial effects to the variability of feed efficiency traits and negligible effects for other performance traits. However, associations between microbial community and complex phenotypes could be identified for almost all traits. These could be considered as future biomarkers for genetic improvement of feed efficiency in pigs.

5.

General discussion

5.1. Overview of the thesis

The main objective of the thesis presented in this document was to investigate how genomic tools applied to the animal and its microbiota can contribute to improve selection for feed efficiency. Using phenotypic and molecular data collected in divergent lines after 10 generations of selection for RFI, we showed through the three result chapters that molecular information acquired on the pigs or their microbiota could be used to complement the existing information and improve selection for feed efficiency. Indeed, as shown in the general introduction of the thesis, feed efficiency is a trait costly to record on all candidates to selection, and highly affected by the production conditions, whereas molecular information could be easier to obtain and could provide complementary information for the selection. Specifically, in the first result chapter we showed that combining phenotypic and molecular information from different related populations can provide a sufficient selection accuracy for such traits, while limiting the genotyping and phenotyping efforts. In the second results chapter, we showed via two complementary approaches, comparing divergent lines and estimating genetic parameters, that the statistical links between feed efficiency and gut microbiota, previously mainly described in phenotypic studies, have some genetic bases that could be exploited for selection. Finally, in the third results chapter, we showed that for feed efficiency traits, the microbiota information can explain a sizable proportion of the trait variance, with limited confounding with the genetic information, and we could identify some of the microbiota components that drive these relationships. Altogether, we can then propose that both pig and microbiota DNA information can provide new information to be used for the genetic improvement of feed efficiency in pigs.

In the following section, we will discuss in more details some of these outcomes, their main limits and potential for application, and then conclude the thesis.

5.2. Efficiency of genomic evaluation using multi-population training sets

As discussed in the introduction section, in pigs, because of the high selection intensities and short generation intervals, the advantage of genomic selection over the traditional pedigreed-based BLUP evaluation in terms of genetic progress essentially depends on an improvement of the accuracy of prediction of the animals BV (Tribout et al., 2013). Achieving a high prediction accuracy with the traditional evaluation is possible by increasing the number of phenotyped animals for the breeding goal and with genomic evaluation by increasing the number of genotyped animals. Therefore, before implementing the genomic selection, expenditures related to the two options should be considered. Nevertheless, the number of affordable phenotypic records in general is highly dependent on the nature of the breeding goal, and is more critical for traits that are sex-limited, late-recorded or expensive to measure. For these types of traits, genomic selection is more promising than selection based on traditional evaluation, because the potential to gain in prediction accuracy is higher (Samore & Fontanesi, 2016). Genomic selection in pigs, unlike in dairy cattle, did not generate structural changes in the selection designs (Tribout et al., 2013) for the moment. The reason is that the selection process in pigs is not based on progeny testing, because most of the economic goal traits are measurable on both male and female animals during their growth, i.e. before selection happens. Therefore, selection candidates mainly have to be reared until the age of realizing their own performances, which is potentially neutralizing the concerns about the breeding costs of the selection candidates. The other reasons concerning the reduced gains of genomic selection efficiency in pigs compared to other species include the lower cost of phenotyping of pig traits compared to the genotyping costs, and the necessity with genomic prediction to maintain continuous phenotyping for traits that are expensive or difficult to measure, for updating the training population for LD changes with time. Thus, the economic benefits of the genomic selection in pigs should be high enough to justify its practical implementation and investing in genotyping. Using low-density marker panels for genotyping and implementing imputation techniques can considerably decrease the total cost of the genomic selection. However, loss of prediction accuracy due to the imputation errors can finally result in loss of genomic selection efficiency. Nevertheless, high accuracy of imputation can be achieved if parent animals are genotyped with the

higher density tool and pedigree information is accurate (Samore & Fontanesi, 2016). In this regard, Carillier-Jacquin et al. (2018) have found sufficient imputation accuracy of low density to medium density SNP panels to use in genomic evaluations in pigs. Another important way to increase the cost efficiency of genomic selection is reducing the number of genotyped animals. This later solution is highly dependent on the heritability of the trait in the breeding goal, as for traits with moderate to high heritability a decrease of the number of animals in the training population would not have a profound effect on the prediction accuracy. In fact, a simulation study conducted by Tribout et al. (2013) showed limited loss of prediction accuracy with 20% or 40% reduction in size of the training population for a trait with heritability of 0.4. Therefore, optimizing the design of training population is an important step that empowers the efficiency of the genomic selection, but should be considered for the different types of traits of the breeding goal. To evaluate the possibility to reduce genotyping costs, and considering that feed efficiency traits are expensive to measure and have moderate heritabilities, in the second chapter we tested 12 scenarios that differed in the design of training populations and comprised animals from 2 related lines. In small size populations, genotypic information for animals from the earlier generations might not be available, which would leave too few animals for constructing a training population allowing high prediction accuracies. Therefore, tested scenarios were set up with consideration of practical aspects of the genomic evaluation when two related lines are available. Results of genomic predictions with these scenarios showed that including a small proportion of animals from a genetically different sub-population in the training set could maintain the prediction accuracy in a standard level, i.e. similar to prediction accuracy based on a homogeneous training population of the same size. Performing genomic selection using across-lines training sets is potentially more feasible than across-breeds training sets, given the presence of pedigree relationships between lines and persistence of similar LD between the sub-populations. In our study, similar accuracy of the scenario comprising animals from the extreme generation of the opposite line in the training set (scenario 6) with the prediction accuracy the routine training set (scenario 1) was mainly due the remaining genomic relationships between animals from the opposite line with the validation population. The other scenarios also had these relationships but maximum, average and minimum values in all

of them were lower than in scenario 6. From a practical point of view, the last scenario was the most economic genomic prediction design for a between-line prediction, as genotypes would benefit for the genomic predictions in both lines. In conclusion, taking advantage of different allelic frequency of a related sub-population to construct the training population is a compromising practical strategy to control the cost of genomic information when initiating genomic prediction. A complete evaluation of the relative costs of phenotyping and genotyping would be needed to better calibrate such strategies, including different types of genotyping tools.

5.3. Dependency of feed efficiency variation to the intestinal microbial composition

As proposed at the beginning of this thesis, a better understanding of the relationships between gut microbiome composition and feed efficiency in pigs could clarify the factors that drive the variability of feed efficiency between animals with respect to the composition of gut microbiome. Indeed, part of the energy produced by gut bacteria is used for their own growth and part becomes available to the host animal (Figure 5-1) (Fetissov, 2017). Interestingly, the indirect energy coming from the bacteria is more efficiently used by the animal than the direct energy extraction from nutrients by the host digestive system (Fetissov, 2017).

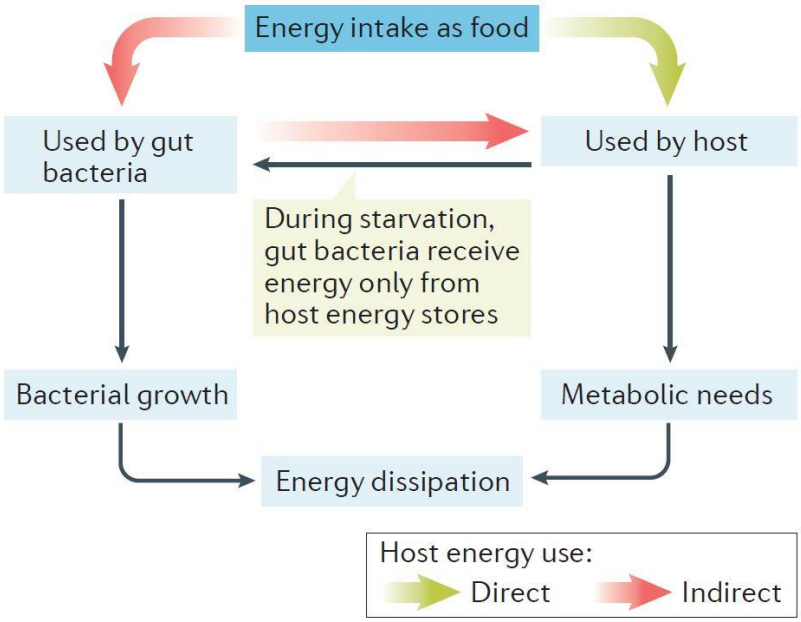


Figure 5-1. Distribution of feed-derived energy between the host and gut bacteria

Source: Fetissov (2017)

In addition to energy, enzymes and metabolites released by bacteria facilitate the digestion of nutrients and fibrous material. Microbiota has also an essential contribution to the appetite and body weight of host animals (Fetissov, 2017; Yang et al., 2018). A study showed that transplantation of gut microbiota from malnourished donor children impaired normal weight gain in recipient mice without causing significant change in food consumption (Blanton et al., 2016). In pigs, Yang et al. (2018) in an investigation on the effect of gut microbiome on host appetite in pigs have found that out of 34 OTUs, 12 OTUs annotated to the *Prevotellaceae* family had positive association with DFI. In their study, some OTUs annotated to the *Ruminococcaceae* and *Lactobacillaceae* families, that are involved in the production of SCFAs and lactic acid, tended to have negative correlations with DFI (Yang et al., 2018). In our study, some genera annotated to the *Ruminococcaceae* and *Lachnospiraceae* families showed significant, and mostly positive, genetic correlations with DFI. The genetic correlations of these genera with RFI and FCR were considerably lower than the corresponding correlation values with DFI. It has been shown that selection for feed efficiency results in reduced appetite in pigs (Eissen et al., 2003; Gilbert et al., 2012). Given these results, reduction of the appetite and increase of

satiety might be favorable in terms of feed efficiency and a selection toward reduction of the genera associated with DFI would not have negative effect on the feed efficiency. Nevertheless, in maternal lines consideration about the milking sows should be taken into account because reduction in the feed intake due to the reduction of appetite may impair the sows metabolism and increase the use of their own body resources during lactation (Gilbert et al., 2012).

In chapter 4, our microbiome-wide association studies showed significant association of few OTUs with feed efficiency and performance traits. These OTUs belonged to the same families in which we found genera with significant genetic correlations with the same traits in chapter 3. Such MWAS approaches, even if not very popular yet in animal studies, seems to be an interesting and complementary tool to target the microbiota components involved in the variability of production traits. If our design had limited power for such analyses, we could consistently point out some OTUs with the two approaches that contribute to the traits variability. These OTUs could actually not be included in earlier variance components estimations with linear mixed models, due to their large number of zeros, so these approaches could be considered as complementary and more exhaustive than those proposed in chapter 3. However, in chapter 4 the main analyses presented corresponded to single-OTU regressions, to overcome the lack of power of simultaneous microbiome-wide estimations. Inspired from the Bayesian alphabet framework developed for genomic predictions, mixture models combining distributions of large and small effects could further contribute to surpass these power limits of the straight microbiome BLUP (M-BLUP) models, and identify the main OTUs contributors to the trait variability. Finally, in addition to the identification to specific OTU or genera related to some traits, the diversity of microbial communities in the gastrointestinal tract of pigs is an indicator of the overall gut health status of animals, and stressful situations can decrease the diversity (Knecht et al., 2020). A lower diversity can cause digestive and finally growth disorders. The alpha-diversity indicators capture the diversity of species within a given sample. The two common alpha diversity metrics used in our study in chapter 3 (i.e. Shannon and Simpson) differed between animals of the LRFI and HRFI lines, and more feed efficient animals had higher gut microbial diversity. This finding not only provides an opportunity for selection programs, but also indicates a

dependency of feed efficiency to the intestinal microbial composition, its diversity and related gut health status.

5.4. Selection of host animals based on microbiome evaluations

Given the relevance of gut microbiome composition to feed efficiency, taking advantage of microbial traits can provide genetic tools to improve feed efficiency. As shown in the previous section, some microbiota components are heritable and genetically associated with production traits, and specifically feed efficiency traits. A first option to use microbiota information for selection for feed efficiency could be to target some few of these components. However, selection for a single microbiota component can cause undesired responses such as decreasing the diversity of microbial communities. Besides, even though we have found moderated heritability for the microbial taxa in the sampling conditions of our design, their abundance may vary in response to changes in age, diet and stressful conditions, which could potentially reduce their robustness as tools for selection proposes. An alternative could be to use more complex profiles associated to traits, such as enterotypes (Mach et al., 2015), but their stability has also been questioned with age and breeding conditions (Le Sciellour et al., 2019). A stronger selection criterion, mentioned in the previous section, could be found in the alpha-diversity metrics, which have moderate heritability and genetic correlations with the studied traits. From a quantitative point of view, because alpha-diversity is a composite measurement, such selection decision could be more robust than a selection based on few selected microbial species or genera. It could capture the ability to maintain functional redundancy in the gut microbiota, rather than to favor some specific components that could disappear in different conditions (Moya & Ferrer, 2016). The diversity and number of microbial communities in the gastrointestinal tract of pigs can also be affected by the destabilization of the intestinal microbiota in early stage of life (Knecht et al., 2020), so it could be recommended to sample pigs after gut microbiota stabilizes when transition occurs (e.g. weaning, dietary changes). However, heritability estimates for the microbial communities indicates a promising stability of the microbial diversity throughout generations.

In addition, special care should be taken to homogenize the overall procedure if microbiota information was used for selection, about the number of samples to initiate it,

but also about the sampling conditions and storage, DBA extraction, sequencing depth and pipeline treatments, and nutritional programs and environmental conditions such as sanitary treatments offered to the pigs. At present, most studies rely on different pipelines and OTU tables can strongly differ in their contents due to post sampling differences in treatments. As a first step, in the present studies OTU were defined for amplicon sequence variants, so any new dataset can be easily combined or compared to ours. However, the taxonomy assignation is itself a field in progress and the difficulty to stabilize long-term options for selection could be an additional constraint for the selection of targeted microbiota components.

As a new field in animal breeding and genetics, results of evaluations based on microbiome data can be used for selection purposes. Thus, another option to use microbiota information for selection is to include it in animal mixed models applied to production traits, as proposed in chapter 4. However, prediction of future phenotypes using microbiome information, relying on microbiability estimations might not be as strong as a genomic prediction because of the GIT location-dependency of the microbiome composition and changes happening with age, diet, and sex (Verschuren et al., 2018; Weishaar et al., 2020). Indeed, the best, i.e. more predictive, microbiome information would certainly result from (combined) samplings at specific GIT locations and at particular age, but this is clearly not affordable for animal selection, for both ethical and economic reasons. Thus, even though selection programs incorporating microbiome information could induce an additional evaluation cost, combined with genetic or genomic evaluations such programs could provide more accuracy for selection for feed efficiency. Careful evaluations of the genetic gains and costs would be needed to decide about these options.

Finally, recent studies proposed to develop the concept of holobiont, that was initially introduced by Margulis and Fester (1991), for selection purposes. A holobiont can be defined as a host animal and all its associated microbiota communities (Simon et al., 2019). Therefore, a selection based on the holobiont concept would involve part of the genome of the host animal that controls a given trait and part of the host genome that control microbiome communities, the so-called “hologenome” (van Vliet & Doebeli, 2019;

Weishaar et al., 2020). The idea of selection based on the hologenome for feed efficiency arises from the partially heritable microbial components that have a substantial effect on feed efficiency, as shown in chapters 3 and 4, and the most recent literature (Weishaar et al., 2020). A selection index using the hologenome would combine a direct genetic effect and a genetic microbial effect in an index. This tends to be similar to the selection of animals based on combinations of their direct genetic value and their maternal genetic value as applied to some economic traits, which incorporates the two different aspects in a selection index. Therefore, a selection index for the hologenome is a host-level selection, organized to improve the host performance traits by retaining both the direct genetic effect and the microbiota effect under genetic control, as proposed by Weishaar et al. (2020) in a two-step strategy. As discussed before, the selection at the microbiome level only can also be considered. However, unlike the hologenome selection that can be optimized for several traits at once, the selection response of a single trait at the microbiome level only would be limited to the improvement of the microbiota composition for the corresponding trait (Weishaar et al., 2020). Our first estimations (chapter 4), as those provided by Khanal et al. (2019), clearly show that genetic and microbiota correlations can differ widely depending on the traits. Here again, a careful evaluation of the improvement of the genetic gains should be run before deciding about the best options for selection.

5.5. Conclusion

In this thesis, the potential of genomic tools applied to the pig and its microbiota to improve selection for feed efficiency has been clarified. We first showed that genomic predictions are feasible for feed efficiency, even when populations are of limited sizes. The next step would be to run an economic assessment to clarify the actual economic potential of this approach. We then showed that the gut microbiota variability contributes to the variability of the production traits, in particular the feed efficiency traits. We identified microbiota components (genera, OTU, α-diversity indexes) which have a genetic background and are associated to different trait levels. Besides, we suggested that accounting for the microbiota information in prediction models could contribute to better prediction accuracy than predictions from the genetic information alone, especially for feed efficiency traits, given the magnitude of the microbiota effects in mixed models. Further studies will be needed to evaluate how genomic information of the host and the microbiota can actually be combined in prediction models to either better predict the breeding values themselves, or even obtain joint predictions of breeding and microbiota values, that would lead to the selection of the hologenome for improved production efficiency.

5.6. References

Blanton, L. V., Charbonneau, M. R., Salih, T., Barratt, M. J., Venkatesh, S., Ilkaveya, O., Subramanian, S., Manary, M. J., Trehan, I., Jorgensen, J. M., Fan, Y. M., Henrissat, B., Leyn, S. A., Rodionov, D. A., Osterman, A. L., Maleta, K. M., Newgard, C. B., Ashorn, P., Dewey, K. G., & Gordon, J. I. (2016). Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *science*, *351*(6275). https://doi.org/10.1126/science.aad3311

Carillier-Jacquin, C., Bouquet, A., Labrune, Y., Brenaut, P., Riquet, J., & Larzul, C. (2018). Using 1K panel in 3 French pig breeds for genomic selection: Accuracy of Imputation and estimation of genomic breeding values using 1K SNP panel, designed for several breeds in French pig populations. 11. World Congress on Genetics Applied to Livestock Production (WCGALP),

Eissen, J. J., Apeldoorn, E. J., Kanis, E., Verstegen, M. W., & de Greef, K. H. (2003). The importance of a high feed intake during lactation of primiparous sows nursing large litters. *Journal of Animal Science*, *81*(3), 594-603. https://doi.org/10.2527/2003.813594x

Fetissov, S. O. (2017). Role of the gut microbiota in host appetite control: bacterial growth to animal feeding behaviour. *Nat Rev Endocrinol*, *13*(1), 11-25. https://doi.org/10.1038/nrendo.2016.150

Gilbert, H., Bidanel, J. P., Billon, Y., Lagant, H., Guillouet, P., Sellier, P., Noblet, J., & Hermesch, S. (2012). Correlated responses in sow appetite, residual feed intake, body composition, and reproduction after divergent selection for residual feed intake in the growing pig. *Journal of Animal Science*, *90*(4), 1097-1108. https://doi.org/10.2527/jas.2011-4515

Khanal, P., Maltecca, C., Schwab, C., Fix, J., & Tiezzi, F. (2019). Microbiability of meat quality and carcass composition traits in swine. *bioRxiv*, 833731.

Knecht, D., Cholewińska, P., Jankowska-Mąkosa, A., & Czyż, K. (2020). Development of swine's digestive tract microbiota and its relation to production indices—a review. *Animals*, *10*(3), 527.

Le Sciellour, M., Zemb, O., Hochu, I., Riquet, J., Gilbert, H., Giorgi, M., Billon, Y., Gourdine, J. L., & Renaudeau, D. (2019). Effect of chronic and acute heat

challenges on fecal microbiota composition, production, and thermoregulation traits in growing pigs1,2. *Journal of Animal Science*, *97*(9), 3845-3858. https://doi.org/10.1093/jas/skz222

Mach, N., Berri, M., Estelle, J., Levenez, F., Lemonnier, G., Denis, C., Leplat, J. J., Chevaleyre, C., Billon, Y., Dore, J., Rogel-Gaillard, C., & Lepage, P. (2015). Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ Microbiol Rep*, *7*(3), 554-569. https://doi.org/10.1111/1758-2229.12285

Margulis, L., & Fester, R. (1991). *Symbiosis as a source of evolutionary innovation: speciation and morphogenesis*. Mit Press.

Moya, A., & Ferrer, M. (2016). Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends in Microbiology*, *24*(5), 402-413. https://doi.org/10.1016/j.tim.2016.02.002

Samore, A. B., & Fontanesi, L. (2016). Genomic selection in pigs: state of the art and perspectives. *Italian Journal of Animal Science*, *15*(2), 211-232. https://doi.org/10.1080/1828051x.2016.1172034

Simon, J.-C., Marchesi, J. R., Mougel, C., & Selosse, M.-A. (2019). Host-microbiota interactions: from holobiont theory to analysis. *Microbiome*, *7*(1), 5. https://doi.org/10.1186/s40168-019-0619-4

Tribout, T., Larzul, C., & Phocas, F. (2013). Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. *Genetics Selection Evolution*, *45*(1), 40. https://doi.org/10.1186/1297-9686-45-40

van Vliet, S., & Doebeli, M. (2019). The role of multilevel selection in host microbiome evolution. *Proc Natl Acad Sci U S A*, *116*(41), 20591-20597. https://doi.org/10.1073/pnas.1909790116

Verschuren, L. M., Calus, M. P., Jansman, A. J., Bergsma, R., Knol, E. F., Gilbert, H., & Zemb, O. (2018). Fecal microbial composition associated with variation in feed efficiency in pigs depends on diet and sex. *Journal of Animal Science*, *96*(4), 1405-1418.

Weishaar, R., Wellmann, R., Camarinha-Silva, A., Rodehutscord, M., & Bennewitz, J. (2020). Selecting the hologenome to breed for an improved feed efficiency in

pigs—A novel selection index. *Journal of Animal Breeding and Genetics*, *137*(1), 14-22.

Yang, H., Yang, M., Fang, S., Huang, X., He, M., Ke, S., Gao, J., Wu, J., Zhou, Y., Fu, H., Chen, C., & Huang, L. (2018). Evaluating the profound effect of gut microbiome on host appetite in pigs. *BMC Microbiology*, *18*(1), 215. https://doi.org/10.1186/s12866-018-1364-8

Curriculum Vitae/Resume

Amir Aliakbari

PhD student
National Research Institute for Agriculture, Food and the Environment (**INRAE**)
Génétique Physiologie et Systèmes d'Elevage (**GenPhySE**)
Address: 24, chemin de Borde-Rouge - Auzeville Tolosane F-31326 Castanet Tolosan
Phone: +33783352983
E-mail: amir.aliakbari@inrae.fr

Education

PhD student in Animal Breeding and Genetics, National Research Institute for Agriculture, Food and the Environment (INRAE), Toulouse-France, Since 2018.
MSc in Animal Breeding and Genetics, Islamic Azad University, Karaj-Iran, 2011- 2013.
BSc in Agricultural Engineering, Animal Science, Islamic Azad University, Maragheh-Iran, 2007 - 2011.

Publications

Journal Papers

Aliakbari, A., Zemb, O., Billon, Y., Barilly, C., Ahn, I., Riquet, J. and H., Gilbert (2021) Genetic relationships between feed efficiency and gut microbiome in pig lines selected for residual feed intake. *Journal of Animal Breeding and Genetics.*

Delpuech, E., **Aliakbari, A.**, Labrune, Y., Fève, K., Billon, Y., Gilbert, H. and J., Riquet (2020) Identification of genomic regions affecting various production traits in pigs divergently selected for feed efficiency, *under review in the Genetics Selection Evolution*.

Aliakbari, A., Delpuech, E., Labrune, Y., Riquet, J. and H., Gilbert (2020) The impact of training on data from genetically related lines on the accuracy of genomic predictions for feed efficiency traits in pigs, *Genetics Selection Evolution*, 52, 57.

David, I., **Aliakbari, A.**, Deru, V., Garreau, H., Gilbert, H., and A., Ricard (2020) Inclusive inheritance for residual feed intake in pigs and rabbits. *Journal of Animal Breeding and Genetics*, 137, 535-544.

Aliakbari, A., Ehsani, A., Vaez Torshizi, R., Løvendahl, P., Esfandyari, H., Jensen, J and P., Sarup (2019) Genetic variance of metabolomic features and their relationship with body weight and body weight gain in Holstein cattle, *J. Anim. Sci*, 97, 3832-3844.

Aliakbari, A., Abbasi, M.A and A., Lavvaf (2015) Study on the influence of genetic and environmental maternal effects on body weight traits in Ghezel sheep breed in rural breeding system, *Animal Science Journal (Pajouhesh & Sazandegi)*, 107, 75-86.

Aliakbari, A., Abbasi, M.A and A., Lavvaf (2014) Maternal effects on average daily gain and Kleiber ratio of Ghezel sheep in rural breeding systems, *Journal of Animal Science Researches (Agricultural Science)*, 25, 109-121.

Conference Presentations

Oral Presentation

Aliakbari, A., Zemb, O., Billon, Y., Barilly, C., Ahn, I., Riquet, J. and H., Gilbert (2020) Genetic relationships between feed efficiency and fecal microbiome in pig lines selected for residual feed intake, 71th *Annual meeting of the European Federation of Animal Science (EAAP)*, online congress.

Aliakbari, A., Delpuech, E., Labrune, Y., Riquet, J. and H., Gilbert (2019) Reliability of genomic predictions for feed efficiency traits based on different pig lines. *70th Annual meeting of the European Federation of Animal Science (EAAP), No. 666, Ghent-Belgium.*

Curriculum Vitae/Resume

Amir Aliakbari

PhD student
National Research Institute for Agriculture, Food and the Environment (**INRAE**)
Génétique Physiologie et Systèmes d'Elevage (**GenPhySE**)
Address: 24, chemin de Borde-Rouge - Auzeville Tolosane F-31326 Castanet Tolosan
Phone: +33783352983
E-mail: amir.aliakbari@inrae.fr

Poster Presentation

David, I., **Aliakbari, A.**, Canario, L., Combes, S., Demars, J., Deru, V., Garreau, H., Gilbert, H., and A., Ricard, (2020). Inclusive inheritance for residual feed intake in pigs and rabbits, 71th *Annual meeting of the European Federation of Animal Science (EAAP)*, online congress.

Aliakbari, A., Delpuech, E., Labrune, Y., Riquet, J. and H., Gilbert (2019) Reliability of the genomic predictions for the feed efficiency related trait based on different pig lines. *Gordon Research Conference, Lucca (Barga)-Italy.*

Aliakbari, A., Abbasi, M.A and A., Lavvaf, (2014) Estimations of genetic parameters of Average Daily Gain and Kleiber Ratio at weaning in Ghezel sheep breed by fitting different animal models, *The 6th Congress on Animal Science of Iran, No. 47, Tabriz, Iran.*

Abbasi, M.A., **Aliakbari**, **A.**, Maghsoudi, A., Pahlavan, R. and F., Gafoori-Kesbi, (2014) Estimates of genetic parameters for early reproductive and composite reproductive traits in Ghezel sheep breed under rural breeding systems, *The 6th Congress on Animal Science of Iran, No. 48, Tabriz, Iran.*

Honors & Awards

Winner of an EAAP scholarship 2020.
Research Grant award of science ministry of Iran as financial support for stay in Denmark (2018)
Student Research Grant award of Division of Animal Sciences, Islamic Azad University of Karaj (2013)
Young Scientist Award by the Iranian Society of Animal Science (2012)
Outstanding Achievement in Research Award, Faculty of Agriculture, Karaj University (2011)
Department of Animal Science Excellent student Award (2010)
The Exceptional talents of Islamic Azad University of Maragheh grant Award (2009)

Work Experience

Guest researcher at the department of QGG, Aarhus University, 2017 – present.
Research Assistant, Tarbiat Modares University (TMU), Iran, 2014 – 2016.

Teaching Experience

Lecturer, Mixed linear models in animal breeding, 2016, Islamic Azad University of Abhar, Iran.

Computer Skills

R
Batch programming
Fortran
Advanced general skills in genomic prediction software (e.g., DMU, WOMBAT, GS3, TM, ASReml, BLUPF90 programs)
General skills in breeding programs software (e.g. QMSim)
General skills in statistical packages (e.g., SAS, SPSS, Minitab)

Curriculum Vitae/Resume

Amir Aliakbari

PhD student
National Research Institute for Agriculture, Food and the Environment (**INRAE**)
Génétique Physiologie et Systèmes d'Elevage (**GenPhySE**)
Address: 24, chemin de Borde-Rouge - Auzeville Tolosane F-31326 Castanet Tolosan
Phone: +33783352983
E-mail: amir.aliakbari@inrae.fr

Fields of Interest

Genomic selection, Quantitative genetics and animal breeding, Mixed models in animal breeding, Whole genome evaluation, Stochastic simulation, Breeding plans for improvement of livestock, Microbiome studies

Attendants Programs

Introduction to graphical models with applications to quantitative genetics and genomics, University of Padova, 2019, (With Guilherme J. M. Rosa and Francisco Peñagaricano).
System biology and gen network inference: application in livestock breeding and genetics, INRAE, 2018, (With Antonio Reverter).
Techniques for Writing and Presenting a Scientific Paper, Aarhus University, 2017, (With Mike Grossman).
Design of Genetic Improvement Programs Course, Aarhus University, 2017, (With Christian Sørensen and Theo Meuwissen).
Workshop on "Reliability and its application". Islamic Azad University of Abhar, 2014.

Languages

Azeri (mother tongue)
Turkish (Proficient)
Persian (Proficient)