Universitat Autònoma de Barcelona

GENOMIC ANALYSIS OF FATTY ACID COMPOSITION AND GUT MICROBIOTA IN PIGS

Daniel Crespo Piazuelo

UNIVERSITAT AUTÒNOMA DE BARCELONA

Departament de Ciència Animal i dels Aliments

Facultat de Veterinària

CENTRE DE RECERCA EN AGRIGENÒMICA

Grup de Recerca de Genòmica Animal

GENOMIC ANALYSIS OF FATTY ACID COMPOSITION AND GUT MICROBIOTA IN PIGS

Daniel Crespo Piazuelo

Doctoral thesis to obtain the PhD degree in Animal Production of the Universitat Autònoma de Barcelona, October 2018

Supervisors

Dr. Josep Maria Folch Albareda

Dr. Maria Ballester Devis

El Dr. Josep Maria Folch Albareda, professor titular del Departament de Ciència Animal i dels Aliments de la Universitat Autònoma de Barcelona (UAB), i la Dra. Maria Ballester Devis, investigadora del Departament de Genètica i Millora Animal del Institut de Recerca i Tecnologia Agroalimentàries (IRTA),

fan constar

que el treball de recerca i la redacció de la memòria de la tesi doctoral titulada '*Genomic analysis of fatty acid composition and gut microbiota in pigs'* han estat realitzats sota la seva direcció per

DANIEL CRESPO PIAZUELO

i certifiquen

que aquest treball s'ha dut a terme al Departament de Ciència Animal i dels Aliments de la Facultat de Veterinària de la Universitat Autònoma de Barcelona i a la unitat de Genòmica Animal del Centre de Recerca en Agrigenòmica,

considerant

que la memòria resultant es apta per optar al grau de Doctor en Producció Animal per la Universitat Autònoma de Barcelona,

i perquè quedi constància, signen aquest document a Bellaterra,

A 29 de octubre de 2018.

Dr. Josep Maria Folch Albareda

Dr. Maria Ballester Devis

Daniel Crespo Piazuelo

Original watercolour painting of the cover by Miguel Ángel Crespo Barberán, 2018.

This work was funded by the *Ministerio de Economía y Competitividad* (MINECO) and the *Fondo Europeo de Desarrollo Regional* (FEDER) projects AGL2014-56369-C2-2-R and AGL2017-82641-R. We acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the "*Severo Ochoa Programme for Centres of Excellence in R&D*" 2016-2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics and the CERCA Programme / *Generalitat de Catalunya*.

Daniel Crespo Piazuelo was funded by a "*Formació i Contractació de Personal Investigador Novell* (FI-DGR)" fellowship provided by the "Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR)" from the *Generalitat de Catalunya* (2015-2018). The predoctoral stay (Short Term Scientific Mission) at the Génétique Animale et Biologie Intégrative (GABI) laboratory of the Institut National de la Recherche Agronomique (INRA) was funded by a COST Action grant of the PiGutNet (www.pigutnet.eu).

“For in a minute there are many days”

Romeo and Juliet (3.5.25)

- William Shakespeare -

Content

Summary/Resumen 11

List of tables 15

List of figures 19

List of publications 25

Related publications by the author 26

Abbreviations 27

1. General introduction 31

1.1 Evolution of the pork production 33

1.2 Pork meat quality 35

1.2.1 Fatty acid composition in adipose tissue and muscle 36

1.3 Fatty acid metabolism 38

1.3.1 De novo fatty acid synthesis 38

1.3.2 Mitochondrial fatty acid β-oxidation 40

1.4 Development of NGS and TGS technologies 42

1.5 Applications of 'omics'-based tools 43

1.6 Genomics in animal breeding 47

1.6.1 Genetic variants used as molecular markers 47

1.6.1.1 Single nucleotide polymorphisms (SNPs) 49

1.6.1.2 Small insertions and deletions (indels) 50

1.6.2 QTL mapping and GWAS 51

1.7 Microbiomics 54

1.7.1 Microbiota profiling through the 16S rRNA gene sequencing 54

1.7.2 Anatomy and physiology of the pig gut 57

1.8 European swine breeds 58

1.9 The IBMAP Consortium 60

2. Objectives 65

3. Articles and studies 69

Paper I. Genome-Wide Association Study for backfat and intramuscular fatty acid composition in three different pig crosses based on the Iberian breed .. 71

Paper II. Indel detection from Whole Genome Sequencing data and association with lipid metabolism in pigs 113

Paper III. Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions 145

Paper IV. Association between the pig genome and its gut microbiota composition 181

4. General discussion 215

4.1 Genetic variants and candidate genes associated with fatty acid metabolism in muscle and adipose tissue 218

4.2 Characterization of the pig gut microbiota along the digestive tract and associations between host genome and microbiota composition 227

4.3 Future perspectives and challenges 235

5. Conclusions 241

6. References 247

7. Annexes 279

7.1 Supplementary material Paper I: 'Genome-Wide Association Study for backfat and intramuscular fatty acid composition in three different pig crosses based on the Iberian breed' 281

7.2 Supplementary material Paper II: 'Indel detection from Whole Genome Sequencing data and association with lipid metabolism in pigs' 296

7.3 Supplementary material Paper III: 'Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions' 300

7.4 Supplementary material Paper IV: 'Association between the pig genome and its gut microbiota composition' 331

Acknowledgements 341

Summary

Pork is one of the most consumed meats worldwide and it is subjected to consumer's preferences. Meat quality is affected by fatty acid (FA) composition in muscle and adipose tissues. Gut microbiota composition can also affect meat quality through the production of metabolites such as short-chain fatty acids. However, the relationship between pig genome and gut microbiota is not fully understood. In the current thesis, several studies have been performed to improve our knowledge about the genetic determinism of FA composition. In addition, the composition of the microbiota along the pig gut and its interaction with the host genome has been also analysed.

Genome-wide association studies (GWAS) were performed among 38,424 single nucleotide polymorphisms (SNPs) and 60 phenotypic traits related to FA composition in backfat and muscle. This analysis was performed in 441 pigs from three different backcrosses: BC1_LD (25% Iberian and 75% Landrace), BC1_PI (25% Iberian and 75% Pietrain), and BC1_DU (25% Iberian and 75% Duroc) belonging to the IBMAP experimental population. Nine regions of the pig genome were associated with twelve backfat traits, while six regions were associated with six intramuscular fat (IMF) traits. A total of 50 candidate genes were proposed to explain the variation in these traits. The most promising candidate genes were *ELOVL3*, *ELOVL6*, *ELOVL7*, *FADS2*, *FASN* and *SCD*. Furthermore, *ELOVL6:c.-394G>A* was the most associated SNP with the percentages of C14:0, C16:0, and C16:1(n-7) in backfat.

With the aim of detecting other variants apart from SNPs, we performed an indel detection with the whole genome sequencing data from seven founders (two Iberian boars and five Landrace sows) of the IBMAP pigs. A total of 1,928,746 indels were found in common among the three programs used (*Dindel*, *SAMtools mpileup,* and *GATK*). Ten indels inside genes related with lipid metabolism (*ASPH*, *C1QTNF12*, *CAPN9*, *CCR7*, *CRP*, *GZMA*, *JMJD1C*, *LYST*, *PEX19* and *SAMD4B*) were genotyped in pigs belonging to the three IBMAP backcrosses, obtaining different allelic frequencies. The *C1QTNF12:c.557_559delCCG* indel was associated with the percentage of eicosadienoic acid (C20:2(n-6)) in IMF.

To describe the microbiota composition along the pig gut, luminal contents of five gut sections (duodenum, jejunum, ileum, and proximal and distal colon) were collected in thirteen Iberian pigs. A total of 1,669 operational taxonomic units (OTUs) grouped in 179 genera were found using the 16S rRNA gene sequencing method. *Lactobacillus*, *Clostridium* and *Prevotella* were the three most abundant genera. Colon samples were more similar among pigs and richer in species than small intestine samples were. The metagenome predictions showed that the energy pathways were different along gut sections.

Finally, to reveal the association between host genome and gut microbiota in pigs, the microbiota composition of the rectum of 285 Iberian × Duroc pigs was obtained using the 16S rRNA gene sequencing method, finding 1,257 OTUs distributed in 101 genera and 18 phyla. Firmicutes and Bacteroidetes were the most abundant phyla. GWAS identified 17 genomic regions of the pig genome associated with the relative abundance of six genera (*Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53* and *Streptococcus*). A total of 38 candidate genes, related with the host defence system and the metabolism of mucopolysaccharides and bile acids, were suggested to be modulators of the gut microbiota composition.

Resumen

La carne de cerdo es una de las carnes más consumidas en el mundo, cuyo valor se ve afectado por su calidad y las preferencias del consumidor. La composición de los ácidos grasos (AGs) en músculo y tejido adiposo modifica la calidad de la carne. Del mismo modo, la microbiota intestinal, a través de la producción de metabolitos como los ácidos grasos volátiles, puede también afectar su calidad. Sin embargo, la relación entre el genoma del cerdo y su microbiota intestinal no está bien estudiada. En la presente tesis se han realizado una serie de trabajos con el fin de profundizar en los mecanismos genéticos implicados en la determinación de la composición de los AGs. Además, se ha estudiado la composición de la microbiota a lo largo del intestino y su interacción con el genoma porcino.

Se realizaron estudios de asociación del genoma completo (GWAS) entre 38.424 *Single Nucleotide Polymorphisms* (SNPs) y 60 caracteres fenotípicos relacionados con la composición de los AGs en músculo y grasa dorsal de 441 cerdos pertenecientes a tres retrocruces de la población experimental IBMAP: BC1_LD (25% Ibérico y 75% Landrace), BC1_PI (25% Ibérico y 75% Pietrain), y BC1_DU (25% Ibérico y 75% Duroc). El GWAS reveló nueve regiones del genoma porcino asociadas con doce caracteres de la grasa dorsal y seis regiones asociadas con seis medidas de la grasa intramuscular. Dentro de estas regiones, se identificaron 50 genes como candidatos funcionales a explicar la variación de estos caracteres. Los genes más relevantes fueron *ELOVL3*, *ELOVL6*, *ELOVL7*, *FADS2*, *FASN* y *SCD*. Además, el polimorfismo *ELOVL6:c.-394G>A* fue el más asociado con los porcentajes de C14:0, C16:0, y C16:1(n-7) en grasa dorsal.

Para estudiar otras variantes genéticas aparte de los SNPs, se detectaron 1.928.746 indels con tres programas (*Dindel*, *SAMtools mpileup,* y *GATK*) mediante los datos de secuenciación del genoma completo de siete fundadores (dos machos Ibéricos y cinco hembras Landrace) del material IBMAP. Se genotiparon diez indels localizados en genes relacionados con el metabolismo lipídico en los 441 cerdos de los tres retrocruces, encontrándose a distintas frecuencias alélicas. En la grasa intramuscular, el indel

C1QTNF12:c.557_559delCCG presentó una asociación significativa con el porcentaje de ácido eicosadienoico (C20:2(n-6)).

Para describir la composición de la microbiota a lo largo del intestino, se recogió el contenido luminal de cinco regiones (duodeno, yeyuno, íleo, colon proximal y distal) de trece cerdos Ibéricos. Posteriormente, mediante el método de amplificación y secuenciación del gen 16S rRNA, se identificaron 1.669 *operational taxonomic units* (OTUs) agrupados en 179 géneros, siendo los más abundantes *Lactobacillus*, *Clostridium* y *Prevotella*. Las muestras de colon eran más ricas en especies y se parecían más entre cerdos que las muestras del intestino delgado. Además, las predicciones funcionales del metagenoma a lo largo del intestino mostraron que sus rutas energéticas eran distintas.

Finalmente, se estudió la asociación entre el genoma del cerdo y su microbiota intestinal. Se obtuvo la composición de la microbiota del recto de 285 cerdos Ibérico × Duroc mediante la amplificación y secuenciación del gen del 16S rRNA, identificándose un total de 1.257 OTUs agrupados en 101 géneros y 18 filos, siendo los filos más abundantes Firmicutes y Bacteroidetes. El GWAS reveló 17 regiones del genoma porcino asociadas con la abundancia relativa de los géneros *Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53* y *Streptococcus*. Dentro de estas regiones, se identificaron 38 genes como candidatos a modular la composición de la microbiota intestinal por su relación con el sistema inmunitario y el metabolismo de los mucopolisacáridos y los ácidos biliares.

List of tables

General introduction

Table 1.1. Morphologic evolution of Landrace breed (adapted from *Producción Animal* Journal as appear in *Manual de diferenciación racial* (Sañudo Astiz, 2008) and AACP (*Asociación Argentina Cabañeros de Porcinos*) data, 2007). 36

Table 1.2. Fatty acids divided by their degree of saturation. 37

Table 1.3. Fatty acids grouped by carbon chain length. 37

Table 1.4. High-throughput SNP genotyping arrays for pigs that are currently available (adapted from Samorè and Fontanesi, 2016 and Blasco and Pena, 2018). 50

Paper I

Paper I. Table 1. Descriptive statistics including mean and SD of intramuscular fat and backfat fatty acid composition and FA indices in the merged dataset of the three backcrosses. 78

Paper I. Table 2. Description of the regions associated with the fatty acid composition in backfat and the candidate genes contained within them. 82

Paper I. Table 3. Description of the regions associated with the fatty acid composition in intramuscular fat and the candidate genes contained within them. 83

Paper II

Paper II. Table 1. Descriptive statistics including mean and SD of fatty acid composition and FA indices in the *Longissimus dorsi* muscle. 120

Paper II. Table 2. Consequences predicted by the *VEP* platform. 129

Paper II. Table 3. Selection of the ten genotyped indels with the alternative allele frequency in the Iberian (Freq. IB) and Landrace (Freq. LD) founders and their consequence predicted by the *VEP* platform. 130

Paper II. Table 4. Genotype frequencies of the nine indels found in each backcross. For each backcross, 148 BC1_DU, 160 BC1_LD and 143 BC1_PI were genotyped. 131

Paper IV

Paper IV. Table 1. Significant genomic regions in the pig genome associated with the relative composition of genera and the candidate genes found within. 194

General discussion

Table 4.1. Strong candidate genes to explain the QTLs for the FA composition of adipose tissue and muscle that were obtained through GWAS and were found in common with the present study in the merged dataset. 219

Annexes: Paper I

Paper I. Supplementary Table S1. List of the significantly associated SNPs within QTL regions for the FA composition in backfat when GWAS were performed in the merged dataset and the predicted consequences of the SNPs. 281

Paper I. Supplementary Table S2. List of the significantly associated SNPs within QTL regions for the FA composition in IMF when GWAS were performed in the merged dataset and the predicted consequences of the SNPs. 289

Paper I. Supplementary Table S3. Summary of the GWAS results for the thirteen SNPs genotyped in the three backcrosses. 292

Paper I. Supplementary Table S4. Frequencies of the thirteen SNPs genotyped in the three backcrosses. 293

Paper I. Supplementary Table S5. Summary of the associated regions found in the three backcrosses when GWAS were performed individually. 295

Annexes: Paper II

Paper II. Table S1. GEMMA output for the suggestive (FDR≤0.1) SNPs found in the GWAS analysis for the log2 normalization of the relative abundance of eicosadienoic acid in the *Longissimus dorsi* muscle of the BC1_PI population. 299

Annexes: Paper III

Paper III. Supplementary Table S1. This table contains the results of the presence/absence analysis at the genus level performed with metagenomeSeq[24]. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which genus is uniquely present in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon. 300

Paper III. Supplementary Table S2. This table contains the results of the differential abundance analysis at the genus level performed with metagenomeSeq[24]. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which genus is more abundant in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and, purple, distal colon. 309

Paper III. Supplementary Table S3. This table contains the results of the differential abundance analysis performed with DESeq2[27] for the KEGG[26] orthologies (KOs) predicted with PICRUSt[25] at the pathway level. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which pathway is more abundant in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon. 314

Annexes: Paper IV

Paper IV. Supplementary Information S1. Full description of the 16S rRNA gene amplification and sequencing. 331

Paper IV. Supplementary Table S1. Means for the relative abundance of the 18 phyla found in rectal contents of 285 pigs. %_Presence indicates the percentage of the pigs where these phyla were found. 332

Paper IV. Supplementary Table S2. Means for the relative abundance of the 101 genera found in rectal contents of 285 pigs. %_Presence indicates the percentage of the pigs where these genera were found. 333

Paper IV. Supplementary Table S3. Description of the 52 significant genera-associated SNPs and their predicted consequences with the Variant Effect Predictor tool (McLaren *et al.*, 2010) (Ensembl release 92). MAF indicates the minor allele frequency. 336

List of figures

General introduction

Figure 1.1. Evolution of worldwide meat production by four livestock species since 1961 (FAOSTAT, 2016). The black line indicates the evolution of the human population as the right vertical axis explains. 33

Figure 1.2. Number of sows by region in the European Union (EUROSTAT, 2013). One dot represents 1,000 sows. 34

Figure 1.3. Fatty acid biosynthesis in plants and animals (adapted from Tazzini, 2013). 39

Figure 1.4. Mitochondrial fatty acid β-oxidation (Sharpe and McKenzie, 2018). 41

Figure 1.5. Generation process of observable phenotypic traits (adapted from Varona, 2017). 44

Figure 1.6. Most common genetic mutations (adapted from U.S. National Library of Medicine, 2018). The reference sequence is mutated by a substitution, an insertion or a deletion modifying the final protein product. 48

Figure 1.7. Quantitative trait locus mapping (adapted from Mackay, 2001). Two divergent populations for a specific trait are crossed to create a F1 population that contains the chromosomes of both progenitors (red and blue bars). F1 individuals are crossed among themselves and produce a F2 population. F2 individuals contain different fractions of the genome of each parental line. Phenotypic traits of each F2 individual are collected and then, associated with molecular markers that cover their entire genome (plot). 52

Figure 1.8. Variable regions of the 16S ribosomal RNA (Yarza *et al.*, 2014). In bold are marked the hypervariable regions from V1 to V9. 56

Figure 1.9. Spatial heterogeneity of the gut microbiota in the human lower gastrointestinal tract (adapted from Donaldson, Lee and Mazmanian, 2016, and Pereira and Berry, 2017). 58

Figure 1.10. Schematic representation of the three IBMAP backcrosses (Landrace, Duroc and Pietran). 60

Paper I

Paper I. Figure 1. Manhattan plot representing the association analysis between the percentages of: C14:0 **(A)**, C16:0 **(B)** and C16:1(n-7) **(C)** in backfat and SNPs distributed along the pig genome, including eight new genotyped polymorphisms for SSC8. *ELOVL6:c.-394G>A* and *PLA2G12A* polymorphisms are included and labelled with a black circle. Red and blue lines indicate those SNPs that are below the genome-wide significance threshold (FDR ≤ 0.05 and FDR ≤ 0.1, respectively). 87

Paper I. Figure 2. Manhattan plot representing the association analysis between the C20:1(n-9)/C20:0 ratio in backfat and SNPs distributed along the pig genome, including five new genotyped polymorphisms for SSC16. The *ELOVL7:c.*1432A>G*, *PIK3R1.P* (rs322671019) and *PIK3R1.E2* (rs331708297) polymorphisms are included and labelled with a black circle. Red and blue lines indicate those SNPs that are below the genome-wide significance threshold (FDR ≤ 0.05 and FDR ≤ 0.1, respectively). 92

Paper I. Figure 3. Comparison between the associated regions along pig chromosomes for backfat FA composition in the merged dataset and in each backcross individually. The shape indicates the backcross or the merged dataset and the colour indicates the phenotypic trait as it is indicated in the legend. 97

Paper I. Figure 4. Comparison between the associated regions along pig chromosomes for intramuscular FA composition in the merged dataset and in each backcross individually. The shape indicates the backcross or the merged dataset and the colour indicates the phenotypic trait as it is indicated in the legend. 98

Paper II

Paper II. Figure 1. Weighted Venn diagram showing the number of indels shared between the three indel detection programs: *Dindel*, *Pindel* and *SAMtools mpileup*. A total of 1,928,746 indels were found in common. 123

Paper II. Figure 2. Distribution of the density of indels across chromosomes calculated as number of indels per Mb. Chromosomes are sorted in increasing order of density value. 124

Paper II. Figure 3. From the total of 1,631,044 indels detected, it is represented the quantity of them according to their length in bp. Insertions are in red and deletions are in blue. 125

Paper II. Figure 4. Manhattan plot representing the GWAS analysis for the relative abundance of eicosadienoic acid in the *Longissimus dorsi* muscle of the BC1_PI population where the *C1QTNF12* indel (blue circle) was suggestive (FDR≤0.1, blue line). 133

Paper II. Figure 5. Multiple sequence alignment based on MULTALIN (Corpet, 1988) of the porcine C1QTNF12 protein sequence with the deletion and the reference sequences of the C1QTNF12 protein in pig, human, cow and mouse. The green arrow points out the deletion. 135

Paper II. Figure 6. JPred4 prediction of the change in the secondary structure of the porcine C1QTNF12 protein when the alanine in the position 186 (A inside the blue rectangle) of the reference sequence (bottom) is deleted (above). Red segments represent alpha helices and green, beta sheets. 135

Paper III

Paper III. Figure 1. a) Stacked area plot of the OTUs grouped by phyla for the 65 samples sorted by intestinal section. **b)** Percentage evolution along the gut of the ten most abundant bacterial genera in the dataset. Segments represent the standard error. 151

Paper III. Figure 2. Descriptive plots made from the OTUs obtained in each sample. **a)** Boxplot of the Shannon α-diversity for the 13 pigs in each intestinal section. **b)** Boxplot of the Whittaker β-diversity for the 13 pigs in each intestinal section. **c)** Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities for the 65 samples of the 13 pigs in each of the 5 intestinal sections (represented by colours). The size of the dot is proportional to the total number of counts in each sample, as represented in the bottom-right rectangle..... 153

Paper III. Figure 3. Five-part Venn diagram for each of the 13 subjects, showing the OTUs shared among the intestinal sections: duodenum (red), jejunum (yellow), ileum (green), proximal colon (blue), and distal colon (purple). The numbers in the diagrams represent how many OTUs were unique in the five intestinal sections or shared between sections as their areas intersect..... 154

Paper III. Figure 4. SparCC[55] PCIT[56] (SPCIT[13]) prokaryotic genus network for partial correlations with an absolute value above 0.65 between log-transformed genus abundances performed in the **a)** duodenum, **b)** jejunum, **c)** ileum, **d)** proximal colon, and **e)** distal colon. The width of the edge represents the degree of the correlation (wider if it is higher), and the colour shows the sign of the correlation: negative (red) and positive (green). The area of the node is proportional to the relative abundance of the prokaryotic genus. The suggested annotation for some genera is enclosed in square brackets.. 155

Paper III. Figure 5. Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities for the metagenome (KEGG[26] orthologies (KOs) counts) predicted through PICRUSt[25] for the 65 samples from the 13 pigs in each of the 5 intestinal sections (represented by colours). In this plot, it can be seen how the predicted functions for the microbiota of the large intestine sections are more similar among individual pigs, while the predicted functions for the microbiota of the small intestine sections have more variation among individual pigs, meaning that large intestinal microbiotas are more likely to perform similar functions. 160

Paper IV

Paper IV. Fig. 1. Stacked area plot of OTUs grouped by phyla for the 285 pig rectal samples. 190

Paper IV. Fig. 2. Plots showing the diversities and dissimilarities measured using the 1,257 OTUs found in rectal contents of 285 pigs. **a)** Boxplot of the Shannon α-diversity. **b)** Boxplot of the Whittaker β-diversity calculated through the Bray-Curtis dissimilarity. **c)** Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities. The size of the dot is proportional to the total number of annotated reads in each sample. 191

Paper IV. Fig. 3. GWAS plot for the relative abundance of the following genera: *Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53*, and *Streptococcus*. The red lines indicate those SNPs that are below the genome-wide significance threshold (FDR ≤ 0.05), while the blue lines indicate those SNPs that are below genome-wide significance threshold (FDR ≤ 0.1). 193

General discussion

Figure 4.1. Summary of the fatty acid biosynthesis in pigs (adapted from Guillou *et al.*, 2010). Highlighted in red are the potential candidate genes to explain the variations in the fatty acid composition of adipose tissue and muscle. Arrows show the elongation and desaturation pathways of fatty acids whereby a potential candidate gene may be affecting a significant trait. 222

Figure 4.2. Factors that can affect the microbiota composition of the pig gut (adapted from Kers *et al.*, 2018). 229

Figure 4.3. Stacked bar plot of the mean relative abundances (%) for the five most abundant phyla along the pig digestive tract. The microbiota composition of the first five regions (duodenum, jejunum, ileum, and proximal and distal colon) belong to the 13 Iberian population, whereas the microbiota composition of the last part, the rectum, belongs to the commercial population that is composed of 285 Iberian x Duroc pigs. Phyla are indicated by the colours in the upper right legend. The scheme of the pig digestive tract has been adapted from Varel, Yen and Kreikemeier, 1995. 232

Figure 4.4. Effects of gut microbiota in weight regulation, insulin resistance and lipid metabolism (Scheithauer *et al.*, 2016). 237

Annexes: Paper III

Paper III. Supplementary Figure S1. Five-part Venn diagram performed for the OTUs shared among sections when combining the datasets from all subjects: duodenum (red), jejunum (yellow), ileum (green), proximal colon (blue), and distal colon (purple). 328

Paper III. Supplementary Figure S2. DESeq2[27] results below a padj≤0.01 cut-off for the four comparisons between each pair of consecutive sections of the KEGG[26] orthologies (KOs) predicted by PICRUSt[25] represented over the KEGG[26] methane metabolism pathway (map00680): **I.** duodenum vs jejunum; **II.** jejunum vs ileum; **III.** ileum vs proximal colon; **IV.** proximal colon vs distal colon. The colour shows which KO was more abundant in that section when comparing the two sections: white, non-significant; red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon. Ambiguous KOs were coloured grey. For clarity of presentation, the methane metabolism pathway was divided into two parts: **a)** The red rectangle shows how the production of methane was more abundant in the distal colon than in the rest of the comparisons. **b)** The two red rectangles represent how the production of acetate was more abundant in the proximal colon than in the ileum and more abundant in the distal colon than in the proximal colon. 329

List of publications

The present thesis is based on the work contained in the list of articles below:

- Paper I. **Crespo-Piazuelo, D.**, Lourdes Criado-Mesas, L., Revilla, R., Castelló, A., Noguera, J. L., Fernández, A. I., Ballester, M., and Folch, J. M. (2018) 'Genome-Wide Association Study for backfat and intramuscular fatty acid composition in three different pig crosses based on the Iberian breed'. *Genetics Selection Evolution* (*submitted*).

- Paper II. **Crespo-Piazuelo, D.**, Lourdes Criado-Mesas, L., Revilla, R., Castelló, A., Fernández, A. I., Folch, J. M., and Ballester, M. (2018) 'Indel detection from Whole Genome Sequencing data and association with lipid metabolism in pigs' (*in preparation*).

- Paper III. **Crespo-Piazuelo, D.**, Estellé, J., Revilla, M., Criado-Mesas, L., Ramayo-Caldas, Y., Óvilo, C., Fernández, A. I., Ballester, M., and Folch, J. M. (2018) 'Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions'. *Scientific Reports*, 8(1), p. 12727.

- Paper IV. **Crespo-Piazuelo, D.**, Migura-Garcia, L., Estellé, J., Criado-Mesas, L., Revilla, M., Castelló, A., Muñoz, M., García-Casco, J. M., Fernández, A. I., Ballester, M., and Folch, J. M. (2018) 'Association between the pig genome and its gut microbiota composition'. *The ISME Journal* (*in revision*).

Related publications by the author

(Not included in the thesis)

- Revilla, M., Puig-Oliveras, A., Castelló, A., **Crespo-Piazuelo, D.**, Paludo, E., Fernández, A. I., Ballester, M., and Folch, J. M. (2017) 'A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits'. *PloS One*, 12(5), p. e0177014.

- Revilla, M., Puig-Oliveras, A., **Crespo-Piazuelo, D.**, Criado-Mesas, L., Castelló, A., Fernández, A. I., Ballester, M., and Folch, J. M. (2018) 'Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions'. *Scientific Reports*, 8(1), p. 2045.

Abbreviations

ACACA	acetyl-CoA carboxylase alpha
ANGPTL4	angiopoietin like 4
BAC	bacterial artificial chromosome
BC1_DU	25% Iberian and 75% Duroc backcross
BC1_LD	25% Iberian and 75% Landrace backcross
BC1_PI	25% Iberian and 75% Pietrain backcross
BLUP	best linear unbiased prediction
bp	base pair
C14:0	myristic acid
C16:0	palmitic acid
C16:1(n-7)	palmitoleic acid
C16:1(n-9)	7-hexadecenoic acid
C18:0	stearic acid
C18:1(n-9)	oleic acid
C18:2(n-6)	linoleic acid
C18:3(n-3)	alpha-linolenic acid
C20:0	arachidic acid
C20:1(n-9)	gondoic acid
C20:2(n-6)	eicosadienoic acid
C20:3(n-3)	eicosatrienoic acid
C20:3(n-6)	dihomo-γ-linolenic acid
CHGA	chromogranin A
ChIP	chromatin immunoprecipitation
ChIP-Seq	ChIP sequencing
CNVs	copy number variants
dbSNP	SNP database
ELOVL3	ELOVL fatty acid elongase 3
ELOVL6	ELOVL fatty acid elongase 6
ELOVL7	ELOVL fatty acid elongase 7
eQTLs	expression quantitative trait *loci*

FA	fatty acid
FABP4	fatty acid binding protein 4
FABP5	fatty acid binding protein 5
FADS2	fatty acid desaturase 2
FAO	fatty acid β-oxidation
FASN	fatty acid synthase
GWAS	genome-wide association studies
IGF2	insulin like growth factor 2
indels	small insertions and deletions
LCAD	long acyl-CoA dehydrogenase
LCFA	long-chain fatty acid
LPL	lipoprotein lipase
MCAD	medium acyl-CoA dehydrogenase
MCFA	medium-chain fatty acid
miRNA	micro RNA
mRNA	messenger RNA
MUFA	monounsaturated fatty acid
NGS	next-generation sequencing
OTUs	operational taxonomic units
pigQTLdb	pig QTL database
PRKAG3	protein kinase AMP-activated non-catalytic subunit gamma 3
PUFA	polyunsaturated fatty acid
QTL	quantitative trait *locus*
RNA-Seq	RNA sequencing
rRNA	ribosomal RNA
RT-qPCR	quantitative reverse transcription
RYR1	ryanodine receptor 1
SCAD	short acyl-CoA dehydrogenase
SCD	stearoyl-CoA desaturase
SCFA	short-chain fatty acid
SFA	saturated fatty acid
SMRT	single-molecule real-time

SNP	single nucleotide polymorphism
STRs	short tandem repeats
TAGs	triacylglycerols
TGS	third-generation sequencing
UROS	uroporphyrinogen III synthase
VFA	volatile fatty acid
VLCAD	very long acyl-CoA dehydrogenase
VLCFA	very long-chain fatty acid
WGS	whole-genome sequencing

1. GENERAL INTRODUCTION

1.1 Evolution of the pork production

Around 330 million tonnes of meat were produced worldwide in 2016. Three species represented 88.3% of the total meat production (FAOSTAT, 2016). Pork was the largest meat produced worldwide (35.82%), followed by chicken meat (32.48%) and beef (20%) (**Figure 1.1**). The three largest pork producing countries were China (45.81%), the United States of America (9.58%) and Germany (4.73%). Spain was ranked fourth representing 3.34% of the total pork production (FAOSTAT, 2016). In addition, most of the Spanish swine livestock was produced in the regions of Aragón and Cataluña (**Figure 1.2**).

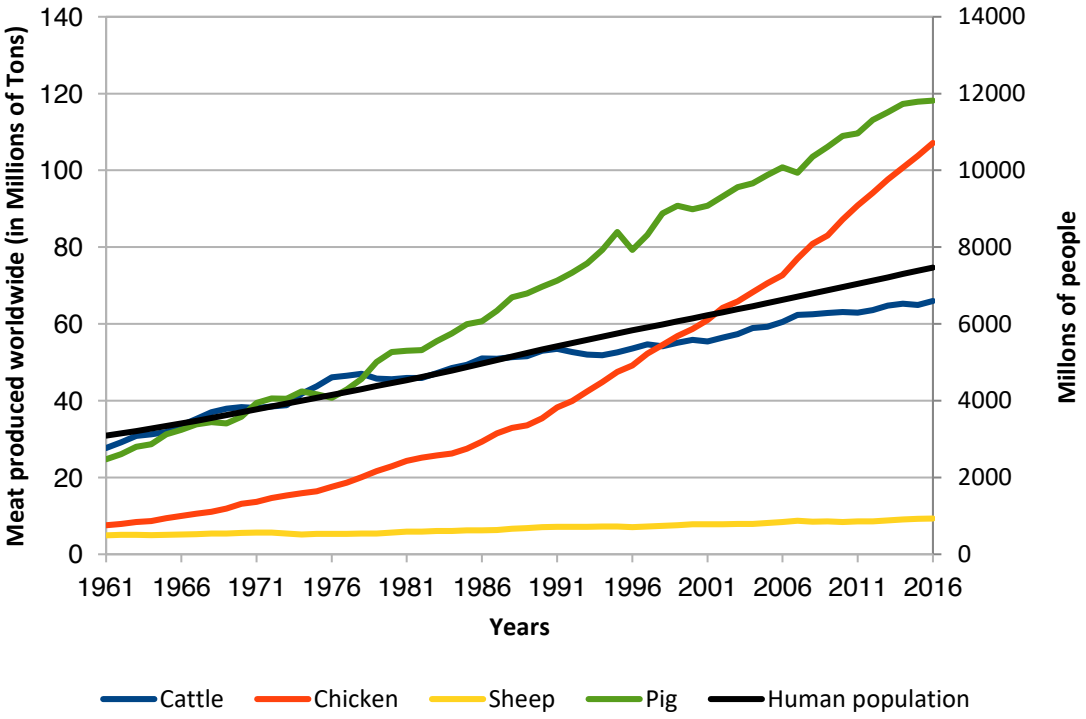


Figure 1.1. Evolution of worldwide meat production by four livestock species since 1961 (FAOSTAT, 2016). The black line indicates the evolution of the human population as the right vertical axis explains.

Five decades ago, meat production was lower, in accordance with meat consumption. Since then, beef production has been increasing in parallel with human population. However, pork and chicken meat productions have grown faster and nowadays people tend to consume more meat. This increased productivity has been achieved due to the optimization of feedstock production and factory farming methods, which have been enhanced by technological and genetic advances. One of the major advances for animal production was the utilization of animal breeding.

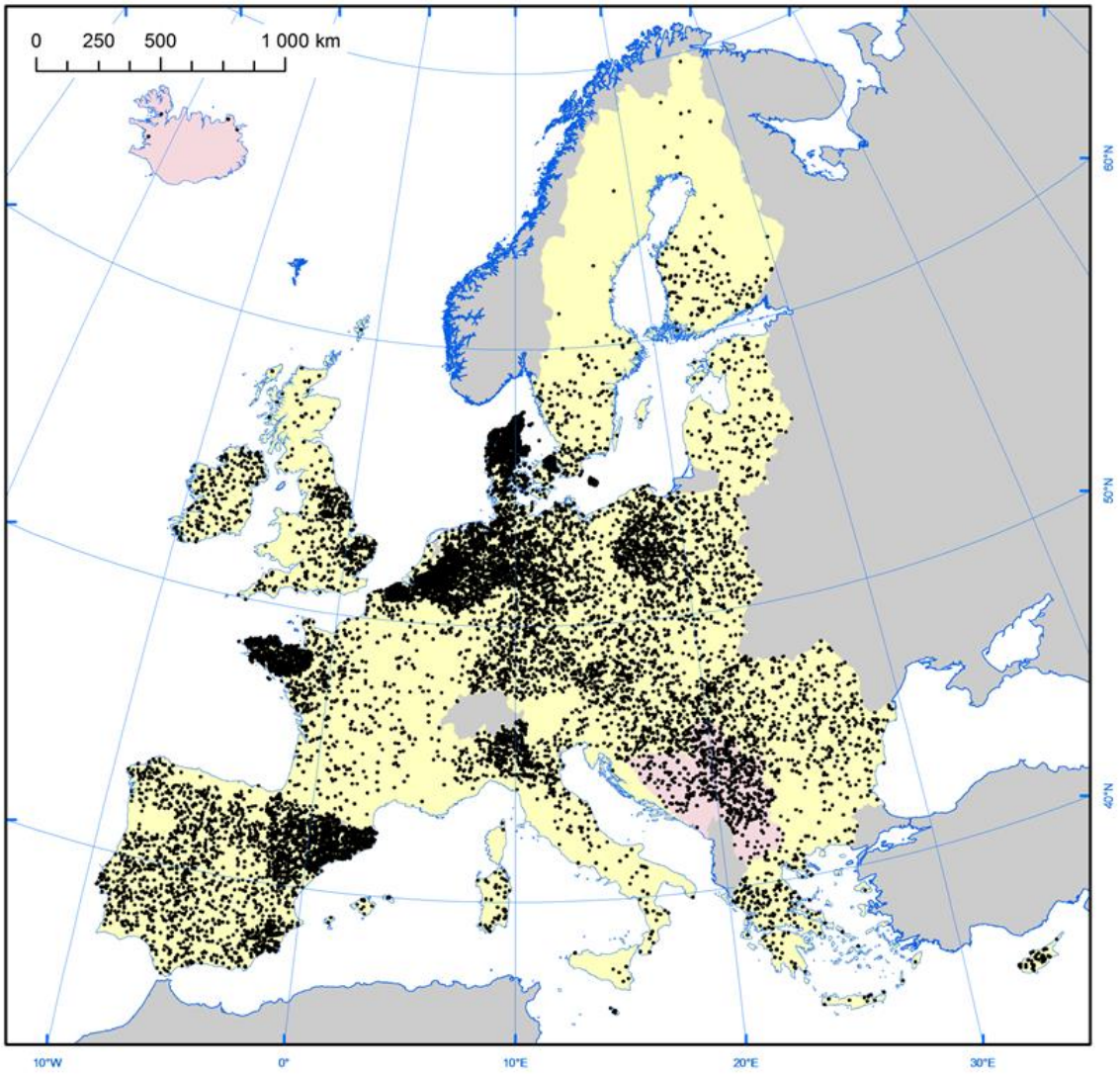


Figure 1.2. Number of sows by region in the European Union (EUROSTAT, 2013). One dot represents 1,000 sows.

1.2 Pork meat quality

Meat quality is a concept difficult to measure and is subjected to consumer preferences. Quality in meat is determined by a list of factors which include freedom from microbiological hazards (food safety), ethical production practices (animal welfare), healthiness and nutritional value (intramuscular lipid content and composition), and sensory aspects (aroma, flavour and taste) (Wood *et al.*, 1999; Webb and O'Neill, 2008; Barendse, 2014). In addition, sensory quality of meat includes raw and cooked appearance and cooked attributes such as texture/tenderness and juiciness (Neethling *et al.*, 2016).

There has been an increasing interest in the healthiness of meat in modern meat consumers, especially in relation to the amount and type of fat (Wood *et al.*, 1999; Webb and O'Neill, 2008). In developed countries, meat is one of the major sources of fat in the diet, especially of saturated fatty acids, which are related with modern life diseases, such as colorectal cancer (Lin *et al.*, 2004) and coronary heart disease (Wood *et al.*, 2004). Therefore, most consumers consider fat as an undesired constituent of meat and label meat as unhealthy (Webb and O'Neill, 2008).

The percentage of meat produced in the carcass has increased through the years due to its economic value, leading to dramatic reductions of fat in pig carcasses, which has been favoured by the consumers' perception of fat as an unhealthy component of meat (Webb and O'Neill, 2008). Thus, pig carcasses have become leaner and longer (Table 1.1) due to the improvement of nutritional requirements and selective breeding (Wood and Whittemore, 2006). However, this increment in leanness has been detrimental for meat quality. The reduction of intramuscular fat content has caused a reduction in tenderness (Wood and Whittemore, 2006). Furthermore, dietary fat is perceived in the oral cavity (Hiraoka *et al.*, 2003) and its reduction influences meat flavour (Wood *et al.*, 1999, 2008). For this reason, meat quality traits have been included in pig breeding programs as consumers are nowadays more interested in healthier and tastier meat products (Wood and Whittemore, 2006).

Table 1.1. Morphologic evolution of Landrace breed (adapted from *Producción Animal* Journal as appear in *Manual de diferenciación racial* (Sañudo Astiz, 2008) and AACP (*Asociación Argentina Cabañeros de Porcinos*) data, 2007).

Trait	1930	1950	1970	2007
Backfat thickness (mm)	40.5	33.4	22.4	13-16.5
Carcass length (cm)	88.9	93.4	96.9	>101

1.2.1 Fatty acid composition in adipose tissue and muscle

Fat and fatty acids (FAs) in adipose tissue and muscle are important for meat quality and its nutritional value. FA composition determine the oxidative stability of muscle and the firmness and oiliness of adipose tissue, which affect flavour and colour of pork (Wood *et al.*, 2008). The incorporation of different oil sources in the diet can modify in a great manner the FA composition of muscle and fat tissues due to the fact that dietary FAs are absorbed intact in the small intestine of monogastric species such as pigs (Wood *et al.*, 1999). Conversely, FA composition of adipose tissue and muscle show moderate to high heritability values in pigs (Cameron, 1990; Cameron and Enser, 1991), manifesting the importance of the genetic component in FA composition.

Certain technological properties of meat are affected by its FA composition. Depending on the number of double bonds (degree of saturation), FAs can be divided in three categories (**Table 1.2**): saturated FAs (SFAs) with no double bonds, monounsaturated FAs (MUFAs) with only one double bond, and polyunsaturated FAs (PUFAs) with two or more double bonds. The ratio of unsaturated to saturated FAs has consequences in the textural and nutritional values of pork and its healthfulness (Wood and Whittemore, 2006; Webb and O'Neill, 2008). SFAs increase cholesterol and low-density lipoprotein (LDL) blood levels, raising the risk of suffering a cardiovascular disease (Mattson and Grundy, 1985; Jiménez-Colmenero *et al.*, 2010). On the other hand, MUFAs and PUFAs, especially omega-3 PUFAs, reduce LDL-cholesterol levels (Bucher *et al.*, 2002). However, while MUFAs confer anti-atherogenic effects and improve meat flavour (Cameron and Enser, 1991), PUFAs are more susceptible to be oxidized, which produce unpleasant organoleptic properties (Wood *et al.*, 2008).

Table 1.2. Fatty acids divided by their degree of saturation.

Name	Acronym	Number of double bonds
Saturated fatty acids	SFAs	No C=C double bonds
Monounsaturated fatty acids	MUFAs	One C=C double bond
Polyunsaturated fatty acids	PUFAs	Two or more C=C double bonds

The degree of saturation and number of carbons that forms the FA chain affect its melting point. The different melting points of FAs affect the colour and the firmness/softness of the fat in meat (Wood *et al.*, 2004). SFAs (e.g. stearic acid: C18:0) have higher melting points than MUFAs of the same length (e.g. oleic acid: C18:1(n-9)) and PUFAs have even lower melting points than MUFAs with the same number of carbons (Berg *et al.*, 2002). In addition, longer FAs will have higher melting points. For example, the melting point of stearic acid (C18:0) is 6.5 degrees higher than that of palmitic acid (C16:0). Fatty acids can be grouped in four categories depending on their carbon chain length (Table 1.3): short-chain FAs (SCFAs), medium-chain FAs (MCFAs), long-chain FAs (LCFAs), and very long-chain FAs (VLCFAs). In this regard, the short chain of SCFAs confers them a high volatility and they are also named volatile FAs (VFAs).

Table 1.3. Fatty acids grouped by carbon chain length.

Name	Acronym	Number of carbons
Short-chain fatty acids	SCFAs	5 or fewer carbons
Medium-chain fatty acids	MCFAs	6 to 12 carbons
Long-chain fatty acids	LCFAs	13 to 21 carbons
Very long-chain fatty acids	VLCFAs	22 or more carbons

1.3 Fatty acid metabolism

Along with proteins, carbohydrates and nucleic acids, lipids are one of the major classes of biomolecules. Lipids are characterized by its solubility in nonpolar solvents and include organic compounds such as FAs, waxes, sterols, fat-soluble vitamins (A, D, E, and K), glycerides (mono-, di-, and tri-), and phospholipids. Lipids are key molecules for energy storage and cell membrane structure (Subramaniam *et al.*, 2011), and establish cellular communications as lipokines (Cao *et al.*, 2008). In addition, lipids participate in the regulation of different biological pathways such as carbohydrate metabolism, neuronal signal transmission and inflammatory response (Subramaniam *et al.*, 2011).

Depending on the nutritional status, FA metabolism works in two directions (lipogenesis and lipolysis). Lipogenesis is produced mainly in the fed state, where excess carbohydrates are used to synthesize FAs in liver and adipose tissue and these FAs are then incorporated into triglycerides for energy storage (Ameer *et al.*, 2014). Lipolysis occurs whenever the body enters in the fasting state, these stored triglycerides are broken down into FAs which are oxidized in mitochondria providing energy for the cells (Frühbeck *et al.*, 2014).

1.3.1 De novo fatty acid synthesis

FA can be obtained through diet or can be *de novo* synthesized in the body. Dietary FAs are absorbed in the small intestine and later incorporated into tissue lipids (Wood *et al.*, 1999). Two FAs are called essential FAs, linoleic acid (C18:2(n-6)) and alpha-linolenic acid (C18:3(n-3)). These two FAs cannot be *de novo* synthesized in mammals, and they must be absorbed through diet (Lagarde *et al.*, 2013). However, once linoleic and alpha-linolenic acids are absorbed, their respective families (omega-6 and omega-3) can be synthesized from these essential FAs (**Figure 1.3**).

In pigs, adipose tissue is the primary site for *de novo* FA synthesis (O'Hea and Leveille, 1969). However, in other species such as humans and rodents, liver is the main site for *de novo* FA synthesis whereas adipose tissue is secondary (Bergen and Mersmann, 2005). Lipogenesis can be also observable in other

tissues such as mammary gland and muscle. Furthermore, the principal precursor of de novo FA synthesis is also different depending on tissues and species. In pigs, glucose is the main source for acetyl-CoA production (O'Hea and Leveille, 1968), while acetate is more used in liver (O'Hea and Leveille, 1969). In other species (e.g., cats), acetate is the principal precursor of acetyl-CoA due to their poor glucose metabolism (Bergen and Mersmann, 2005).

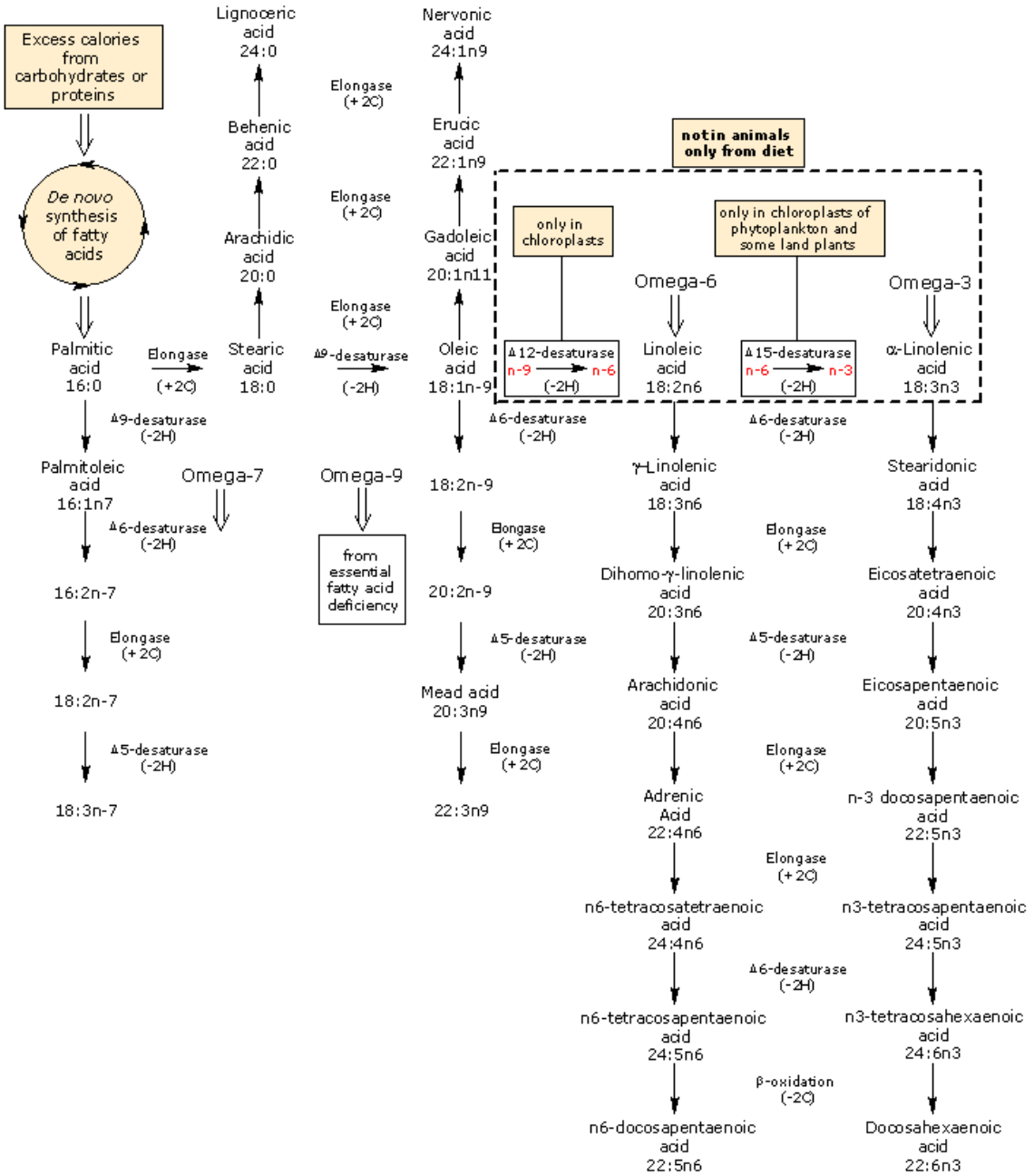


Figure 1.3. Fatty acid biosynthesis in plants and animals (adapted from Tazzini, 2013).

The acetyl-CoA obtained by glucose or acetate users is carboxylated by the acetyl-CoA carboxylase (ACACA) enzyme to generate malonyl-CoA (Bergen and Mersmann, 2005). Then, acetyl-CoA and seven malonyl-CoA molecules are converted into palmitic acid (C16:0) by the fatty acid synthase (FASN), which increases the number of carbons of the acetyl-CoA molecule through successive incorporations of two carbons of the malonyl-CoA group in the presence of NADPH (Wakil *et al.*, 1983). Afterwards dietary FAs, including essential FAs, and *de novo* synthesized FAs suffer different cycles of elongations and desaturations (Figure 1.3). FA elongases increase the length of FAs adding two carbons of the malonyl-CoA group, whereas each acyl-CoA desaturase introduces a double bond in a specific position resulting in the formation of different MUFAs and PUFAs (Guillou *et al.*, 2010).

1.3.2 Mitochondrial fatty acid β-oxidation

In the fasting stage, FAs are used to provide energy through mitochondrial fatty acid *β*-oxidation (FAO) when glucose is not available. In this situation of energy shortage, lipolysis of white adipose tissue triacylglycerols (TAGs) stored in lipid droplets releases FAs and glycerol, which can later be used by other tissues (Ahmadian *et al.*, 2010). In addition, heart, liver and skeletal muscle are also tissues with a high FAO activity (Wajner and Amaral, 2015).

In the cytosol, free FAs are converted into fatty acyl-CoA esters by acyl-CoA synthetases which are later introduced into the mitochondrion through the carnitine shuttle system (Figure 1.4). Inside the mitochondrion, fatty acyl-CoA chains are shortened until one acetyl-CoA molecule is obtained through several cycles of four steps (dehydrogenation, hydration, a second dehydrogenation and thiolysis) (Eaton *et al.*, 1996). After each cycle, a fatty acyl-CoA is shortened by two carbons and acetyl-CoA, NADH and $FADH_2$ are generated (Sharpe and McKenzie, 2018). The first dehydrogenation step has four enzymes with different chain-length specificities which compete for acyl-CoA substrates (van Eunen *et al.*, 2016). These four enzymes are called short-, medium-, long-, and very long-chain acyl-CoA dehydrogenases (SCAD, MCAD, LCAD, VLCAD). Located in the mitochondrial matrix, SCAD, MCAD, and LCAD are active with C4-C6, C4-C12,

and C8-C20 chain-length fatty acyl-CoAs respectively, while VLCAD is the only one located in the inner mitochondrial membrane and is active with C12-C24 chain-length fatty acyl-CoAs (Eaton *et al.*, 1996). The number of double bonds in FAs are also important for the FAO process, PUFAs enter at low rates in the mitochondrion and can act as FAO inhibitors (Osmundsen and Bjørnstad, 1985; Gavino and Gavino, 1991). Furthermore, some PUFAs are difficult to *β*-oxidize due to the presence of *cis*-5 double bonds or *cis* double bonds at even-numbered carbon atoms (Eaton *et al.*, 1996).

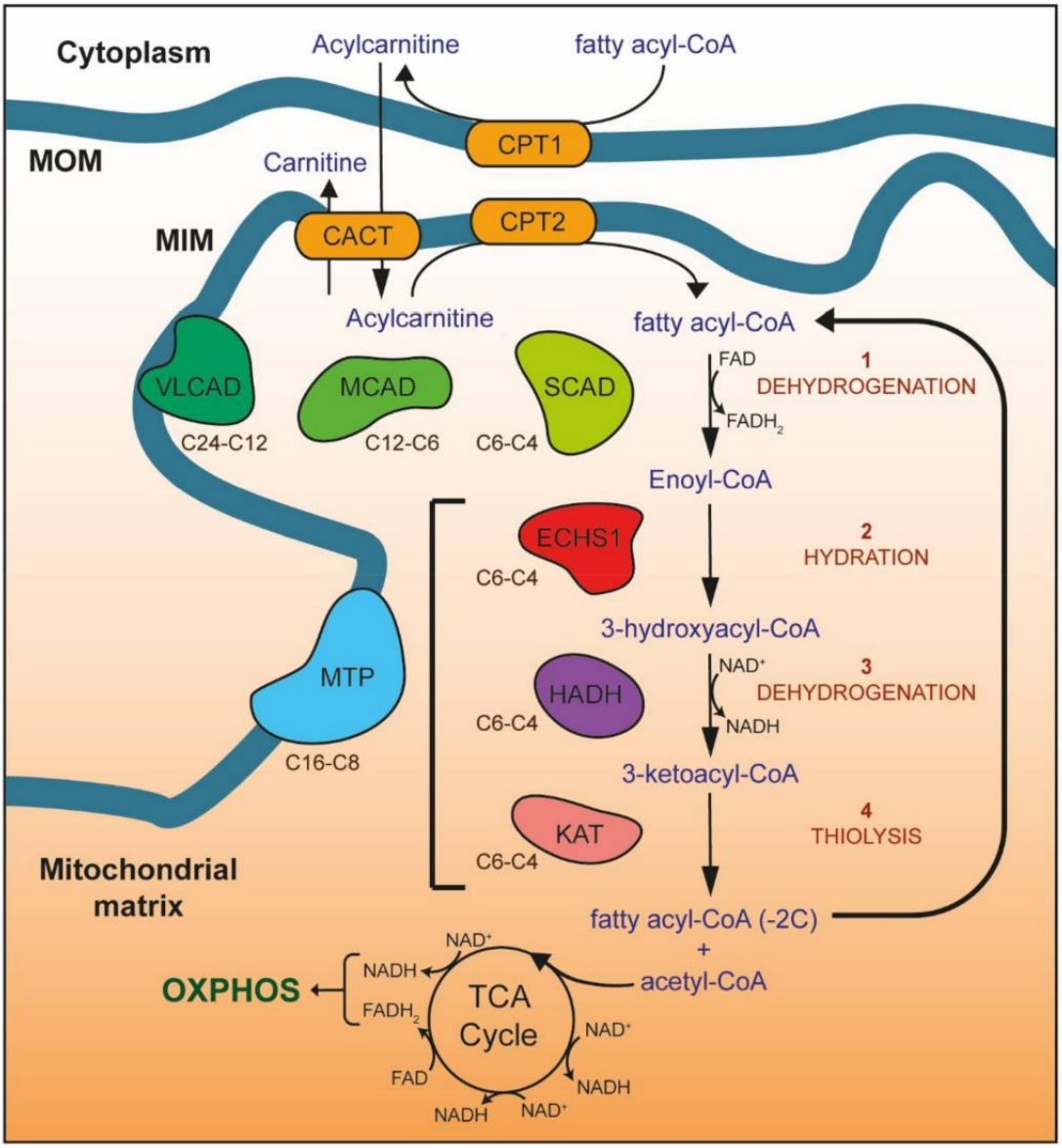


Figure 1.4. Mitochondrial fatty acid β-oxidation (Sharpe and McKenzie, 2018).

1.4 Development of NGS and TGS technologies

In 1977, Sanger and collaborators published a method to sequence DNA by chain terminating inhibitors (Sanger *et al.*, 1977). Since then and for 30 years, Sanger sequencing was the prevalent method to sequence genes and genomes, culminating with the completion of the first human genome sequence (International Human Genome Sequencing Consortium, 2004). Thereafter, a second generation of sequencing technologies were developed and commercialized. These new methods were called next-generation sequencing (NGS) in contrast to the first-generation Sanger method. Three major improvements were shared among NGS technologies (van Dijk *et al.*, 2014):

- They do not require bacterial cloning of DNA fragments. NGS libraries are prepared in a cell free system instead.
- They can run thousands-to-many-millions of sequencing reactions in parallel.
- Not electrophoresis is needed for base interrogation as the sequencing output is directly detected.

All these advances generated a massive amount of reads and entire genomes could be sequenced at a shorter time. Nevertheless, the disadvantage of the relatively short reads produced by NGS technologies provoked that new alignment algorithms were developed to perform genome assembly (van Dijk *et al.*, 2014).

With the commercialization of the first NGS method in 2005 by *454 Life Sciences* (now *Roche*) (Margulies *et al.*, 2005), five platforms have controlled the NGS market since then: 454 pyrosequencing (*Roche*), Illumina (formerly *Solexa*), Sequencing by Oligo Ligation Detection (SOLiD, *Thermo Fisher*), Ion Torrent (*Thermo Fisher*), and PacBio (*Pacific Biosciences*). However, this last platform (PacBio), released in 2011 its first instrument (PacBio RS) which was able to outperform the short-length reads drawback of the rest of the NGS technologies producing several thousands of long reads (Eid *et al.*, 2009). PacBio named its method ‘single-molecule real-time’ (SMRT) sequencing and marked the beginning of third-generation sequencing (TGS) also known as long-read sequencing. Three years later, Oxford Nanopore Technologies (ONT) presented

another TGS technology, a pocket-sized nanopore sequencer called MinION (Jain *et al.*, 2015). Aside from the higher length of reads, TGS are single-molecule sequencing technologies which are performed in real time, in contrast to NGS technologies which need to pause the sequencing reaction after each base incorporation (Schadt *et al.*, 2010). Nevertheless, TGS technologies need to be improved because they introduce a high error rate at single pass (~15%) and are expensive nowadays (van Dijk *et al.*, 2018).

1.5 Applications of 'omics'-based tools

Animal breeding allows the selection of animals regarding the genetic value that has been assigned to each one using different methods. Since the 30's, quantitative genetics and statistics have been used to select animals, but the highest improvement in animal breeding arose in the 70's, with the development of calculation power in computers and some statistical methods such as BLUP (best linear unbiased prediction) (Varona, 2017). In the past decade, these statistical methods have been improved thanks to high-throughput genotyping platforms, which cover the entire genome with markers, allowing the calculation of a genetic kinship matrix in a population and conduct selective breeding for specific traits. Nowadays, genome selection can be performed using whole-genome sequencing (WGS) data, including all the sequenced genomes of the individuals of a population. However, the incorporation of WGS data may not produce a higher increment in the selection response over high-throughput genotyping platforms (Pérez-Enciso *et al.*, 2015).

Phenotypes, such as carcass length or weight, are observable production traits that can be selected in animal breeding. However, the biological processes involved in the generation of these phenotypes are complex and the relationship between genome and phenotype is not direct (**Figure 1.5**). Moreover, the genome is modified by epigenetic marks (epigenome), which are involved in the transcription of DNA into RNA (transcriptome). Messenger RNAs are translated into a set of different proteins (proteome), which can suffer further modifications and produce certain metabolites (metabolome) that will determine a phenotypic

trait. Furthermore, the environment influences the phenotype and interacts with the previous intermediate levels of gene expression. The microbiome also interacts with the environment, producing metabolites which can affect the phenotype and host's gene expression.

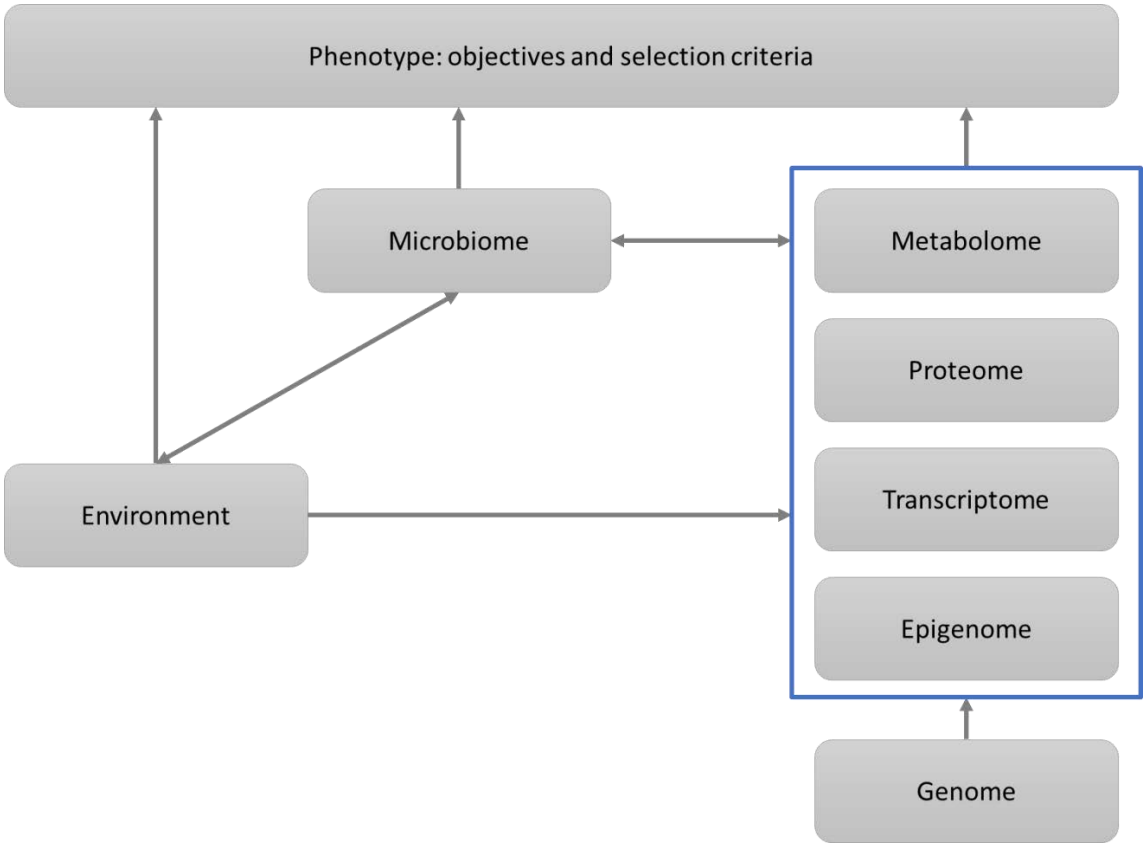


Figure 1.5. Generation process of observable phenotypic traits (adapted from Varona, 2017).

A variety of high-throughput experimental technologies have been developed to unravel the complexity of living cells (Palsson, 2002). The scientific fields that use these new technologies to study the levels from genome to phenotype are called 'omics'.

Genomics. NGS technologies have reduced the time and cost of studying the genome of an individual (Ohashi *et al.*, 2015). DNA is fragmented into millions of sequences that are read through a great number of reactions in parallel. In this context, WGS is the method used to sequence the complete genome of an individual (Mardis, 2008). In addition, whole-exome sequencing is another NGS

methodology focused in the study of the exon regions of the genome (Ohashi *et al.*, 2015). Genomics also analyses the structure and function of a genome, and how it has evolved over time.

Epigenomics. The genome of an individual does not have many variations along its whole life, but the expression of certain parts of the genome is influenced by DNA accessibility and chromatin structure. This epigenetic information is stored as chemical modifications in cytosine bases and histone proteins that affect genome expression across different developmental stages, tissues, and disease states (Bird, 2002; Goll and Bestor, 2005; Margueron *et al.*, 2005). In addition, these chemical modifications can be inherited (Bernstein *et al.*, 2007) and thus, selected (Goddard and Whitelaw, 2014). The application of NGS technologies to chromatin immunoprecipitation (ChIP) assays resulted in a new technique, ChIP sequencing (ChIP-Seq), which allows the identification of genome-wide DNA binding sites for transcription factors and other proteins (Pellegrini and Ferrari, 2012).

Transcriptomics. DNA sequences can be transcribed into a variety of RNAs with different functions, such as messenger RNAs (mRNAs), ribosomal RNAs (rRNAs), or microRNAs (miRNAs). The collection of all the transcribed RNAs is called transcriptome. Expression levels of mRNAs are usually correlated with the amount of protein in which is translated. However, this relationship is not trivial and can be modified by regulatory elements, miRNAs or protein autophagy (Liu *et al.*, 2016). In addition, RNA expression levels can be considered as another trait and be subjected to selection as well. In the same manner that WGS has contributed to the study of the genome, RNA sequencing (RNA-Seq) has contributed to the study of the transcriptome (Ohashi *et al.*, 2015).

Proteomics. All the set of proteins that form part of a genome, cell, tissue or organism in a given time is defined as proteome. Proteomics characterizes the structure, functions, interactions and modifications of these set of proteins at any stage (Aslam *et al.*, 2017). The aminoacidic sequences of mixtures of proteins and their post-translational modifications can be identified through tandem mass spectra and later be matched against sequence databases (Eng *et al.*, 1994).

Metabolomics. Metabolites are small molecules that suffer chemical transformations within the cells of organisms. Therefore, metabolites provide direct signatures of the biochemical activity of the organism and are easier to correlate with phenotypic traits (Patti *et al.*, 2012). High-throughput techniques such as mass spectrometry or nuclear magnetic resonance can be applied to study, in a large-scale manner, the metabolome (Dettmer *et al.*, 2007).

Microbiomics. The microbiota term was firstly defined as the ecological community of commensal, symbiotic and pathogenic microorganisms present in a defined environment (Lederberg and McCray, 2001; Marchesi and Ravel, 2015). From a global perspective, the microbiota (bacteria, archaea, lower and higher eurkaryotes, and viruses), their genomes (i.e., metagenome) and the surrounding environmental conditions form an entire habitat defined as microbiome (Marchesi and Ravel, 2015). Therefore, microbiomics is the scientific field involved in the study of the microbiome. NGS have also been applied to microbiomics, whole-metagenome shotgun sequencing method (Tyson *et al.*, 2004; Venter *et al.*, 2004) allows the sequencing of the complete genome of all the microorganisms that are present in a sample. Cheaper strategies have focused on the amplification and sequencing of hypervariable regions of a gene (e.g.; the 16S rRNA gene), which can be used to differentiate the microorganisms present in a sample (Gray *et al.*, 1984).

Phenomics. Despite phenomics is defined as the acquisition of high-dimensional phenotypic data on an organism-wide scale, phenotypes are not as stable as the genome (Houle *et al.*, 2010). Therefore, phenomics tends to collect as much data as possible in a great number of individuals, characterizing a huge number of phenotypic traits in an easier way and in a reduced time. For instance, lean content and carcass quality are measured with ultrasounds (Brøndum *et al.*, 1998), and through digital images, pig leg conformation can be evaluated (Stock *et al.*, 2017). Nowadays, there are farming systems that integrate individual information of water and feed consumption, growth, environment conditions and even respiratory diseases, which can be used for selective breeding (Varona, 2017).

1.6 Genomics in animal breeding

The application of genomics in animal breeding was conceived with the objective to perform marker-assisted selection. The first efforts to map the genomes of livestock species started in the 1990s, in order to identify markers linked to quantitative trait *loci* (Haley *et al.*, 1990). The *Swine Genome Sequencing Consortium* (SGSC) was formed in 2003 with the aim of coordinate the sequencing of the pig genome (Schook *et al.*, 2005). This project was achieved by the combination of two different strategies, a first hierarchical shotgun sequencing of the bacterial artificial chromosome (BAC) clones (Humphray *et al.*, 2007), which was later improved using *Illumina* NGS data obtained through whole-genome shotgun sequencing (Archibald *et al.*, 2010). After these efforts, the *Sscrofa10.2* assembly of the pig genome sequence was published in 2012 (Groenen *et al.*, 2012). The *Sscrofa10.2* assembly was based on the genome of a female Duroc pig and since its publication, the genome of hundreds of pigs of several breeds have been re-sequenced and more than 350 complete genomes are publicly available (Groenen, 2016). Recently, the *Sscrofa10.2* version was improved thanks to the publication of the *Sscrofa11.1* assembly in 2017. The sequence data used to construct this new assembly was obtained through TGS technologies (PacBio RSII long reads), generating a 65x genome coverage over a total sequence length of 2.5 Gb.

All these improvements in the annotation of the pig genome have been useful for other NGS methods that require a complete genome reference, such as RNA-Seq (transcriptomics) and ChIP-Seq (epigenetics). Furthermore, the publication of a reference genome allowed the identification of gene markers, which are now applied to marker-assisted selection.

1.6.1 Genetic variants used as molecular markers

A molecular marker or genetic marker is a DNA sequence with an identifiable physical location (locus) in a chromosome and whose inheritance can be traced (National Institutes of Health, 2018). Molecular markers are usually variations in DNA sequences that allow the differentiation of individuals in a population.

Furthermore, molecular markers close to each other in a chromosome tend to be inherited together, which is useful to trace phenotypic differences among individuals. In this respect, molecular markers are fundamental to construct genomic maps, detect quantitative trait loci and perform association studies of phenotypic traits.

A mutation is defined as any change in the sequence of a nucleotide or in the organization of the DNA. Mutations can arise from DNA replication errors made during cell division, infection by virus, or exposure to mutagens or ionizing radiations (Nussbaum *et al.*, 2008). Genetic mutations can be produced by base pair substitutions, insertions, deletions, duplications, inversions or fusions of DNA sequences. The two most common mutations are single nucleotide polymorphisms and small insertions and deletions (Weber *et al.*, 2002) (Figure 1.6).

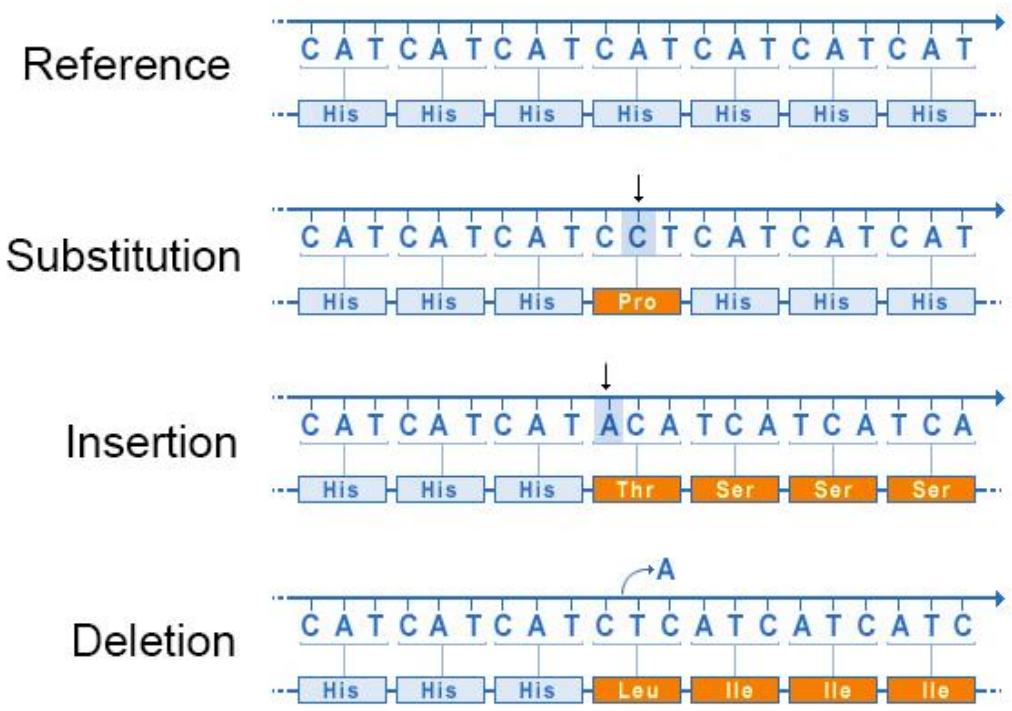


Figure 1.6. Most common genetic mutations (adapted from U.S. National Library of Medicine, 2018). The reference sequence is mutated by a substitution, an insertion or a deletion modifying the final protein product.

1.6.1.1 Single nucleotide polymorphisms (SNPs)

The substitution of one nucleotide base for another in the genome is defined as a single nucleotide polymorphism (SNP) and are the best studied mutations. SNPs in coding regions are defined as synonymous substitutions when the mutation does not change the amino acid sequence of the protein, while a nonsynonymous substitution change one amino acid for another (missense) or introduce a premature stop codon that truncates the translation (nonsense).

The creation of a porcine reference assembly and the development of NGS technologies allowed the massive detection of SNPs in pig genomes (Ramos *et al.*, 2009). Massive SNP detection is usually performed through the combination of whole-genome sequencing (WGS) with mapping pipelines and variant calling software. Nowadays, the last available build of the *Single Nucleotide Polymorphism database* (dbSNP) contains more than 67 million porcine SNPs (Build 150: https://www.ncbi.nlm.nih.gov/projects/SNP/; released on February 2017).

Massive SNP detection led to the commercialization of high-throughput genotyping arrays (a collection of different SNPs distributed along the entire pig genome). High-throughput genotyping arrays are powerful genomic tools for animal breeding that offer a wide range of applications, such as the study of marker-trait association, evaluation of pure lines, construction of kinship matrices and the analysis of genome-wide associations and selective sweeps. These SNP arrays can estimate the genetic value of animals at a young age and thus, increase the rate of genetic improvement by reducing the generation interval (Meuwissen *et al.*, 2001).

The first high-throughput genotyping array was commercialized by Illumina in 2008, although the sequencing of the pig genome had not yet been completed (Ramos *et al.*, 2009). This first array, *PorcineSNP60 v1 BeadChip*, contained more than 60,000 SNPs and was later improved in 2012 with a second version (Table 1.4). The most recent high-throughput SNP genotyping array included more than 650,000 SNPs with an average interval between SNPs of 3.34 kb (Groenen, 2015). However, an increment in the number of SNPs does not lead to a much higher accuracy of prediction for breeding values (Blasco and Pena,

2018). In dairy cattle, when using 500,000 markers instead of 50,000 only a 1.6% of accuracy was gained (VanRaden *et al.*, 2011). Nonetheless, this increment in the number of SNPs would be more indicated to perform fine mapping of quantitative trait loci.

Table 1.4. High-throughput SNP genotyping arrays for pigs that are currently available (adapted from Samorè and Fontanesi, 2016 and Blasco and Pena, 2018).

Chip name	No. of SNPs	Gap spacing[a] (kb)	Company	Technology
PorcineSNP60 v2 BeadChip	64,232	43.4	*Illumina*	*Illumina Infinium®* chemistry
GeneSeek® Genomic Profiler™ for Porcine HD (GGP-Porcine HD)	70,231	43.0	*GeneSeek /Neogen*	*Illumina Infinium®* chemistry
Axiom® Porcine Genotyping Array	658,692	3.34	*Affymetrix*	Axiom assay

[a] Average interval between SNPs.

1.6.1.2 Small insertions and deletions (indels)

In contrast to SNPs and larger structural variants, which have received considerable attention, small indels have been less studied. However, the genome-wide ratio of indels to SNPs has been estimated as 1 indel for every 5.3 SNPs (Mills *et al.*, 2011). Depending on the number of alleles, indels can be divided into two categories: those with two alleles (diallelic) and those with multiple alleles (multiallelic). Most of the multiallelic indels are based on short tandem repeats (STRs), also called 'microsatellites', which were the predominant type of molecular markers used in genetic studies since 1990 until the development of high-throughput genotyping arrays (Weber *et al.*, 2002). Conversely, diallelic indels have received less attention. The difference between alleles in diallelic indels range from one base to large chromosomal regions on the order of megabases (Lupski *et al.*, 1996; Mills *et al.*, 2011). However, diallelic indels that differ only few bases between their alleles are the most common

(Weber *et al.*, 2002). These small indels can be easily detected by NGS methods, but the detection of larger structural variations, such as long indels and copy number variants (CNVs), were more challenging to characterize until the appearance of long-read sequencing (van Dijk *et al.*, 2018).

In the coding regions of the genome, the ratio of indels to SNPs decreases to 1 indel for every 21.7 SNPs (Mills *et al.*, 2011). While SNPs usually produce synonymous changes with slight or no impact on gene function, indels generate major changes in coding exons. In-frame mutations are produced when the number of bases of the indel is divisible by three, whereas if the number of bases is not divisible by three, the reading frame is modified and the codons after the mutation are translated into different amino acids (frameshift mutation). Therefore, indels that map to coding exons are more prone to suffer strong purifying selection and be eliminated much more frequently than SNPs.

1.6.2 QTL mapping and GWAS

Quantitative trait *locus* (QTL) analyses are statistical methods that correlate genomic regions with phenotypic traits (Miles and Wayne, 2008). The genotypic data used to search QTLs are usually molecular markers, such as SNPs or microsatellites, which cover the entire genome. Two molecular markers in the same chromosome tend to be inherited together as nearer they are, because the chance of recombination between them is lower (Figure 1.7). QTL mapping is a powerful method to identify genomic regions that co-segregate with a given trait in intercrossed populations using markers to perform a linkage analysis (Goddard and Hayes, 2009). However, its power is restricted to the allelic diversity that segregates between the parents of the population and the amount of recombination (Korte and Farlow, 2013). High-throughput SNP genotyping arrays increased the number of available genetic markers and as a result, the genome-wide association study (GWAS) approach was developed. GWAS outstands QTL mapping because it uses the recombination events after mutations occur (Meuwissen and Goddard, 2000; Goddard and Hayes, 2009). In addition, thanks to the high density of mapped SNPs and the substitution of pedigree matrices for genetic kinship matrices, GWAS increased the accuracy of QTL analysis,

especially if multiple breeds were considered (Goddard and Hayes, 2009). Nonetheless, GWAS are dependent on the phenotypic variance within the population explained by the marker and are affected by rare variants and small effect size (Korte and Farlow, 2013).

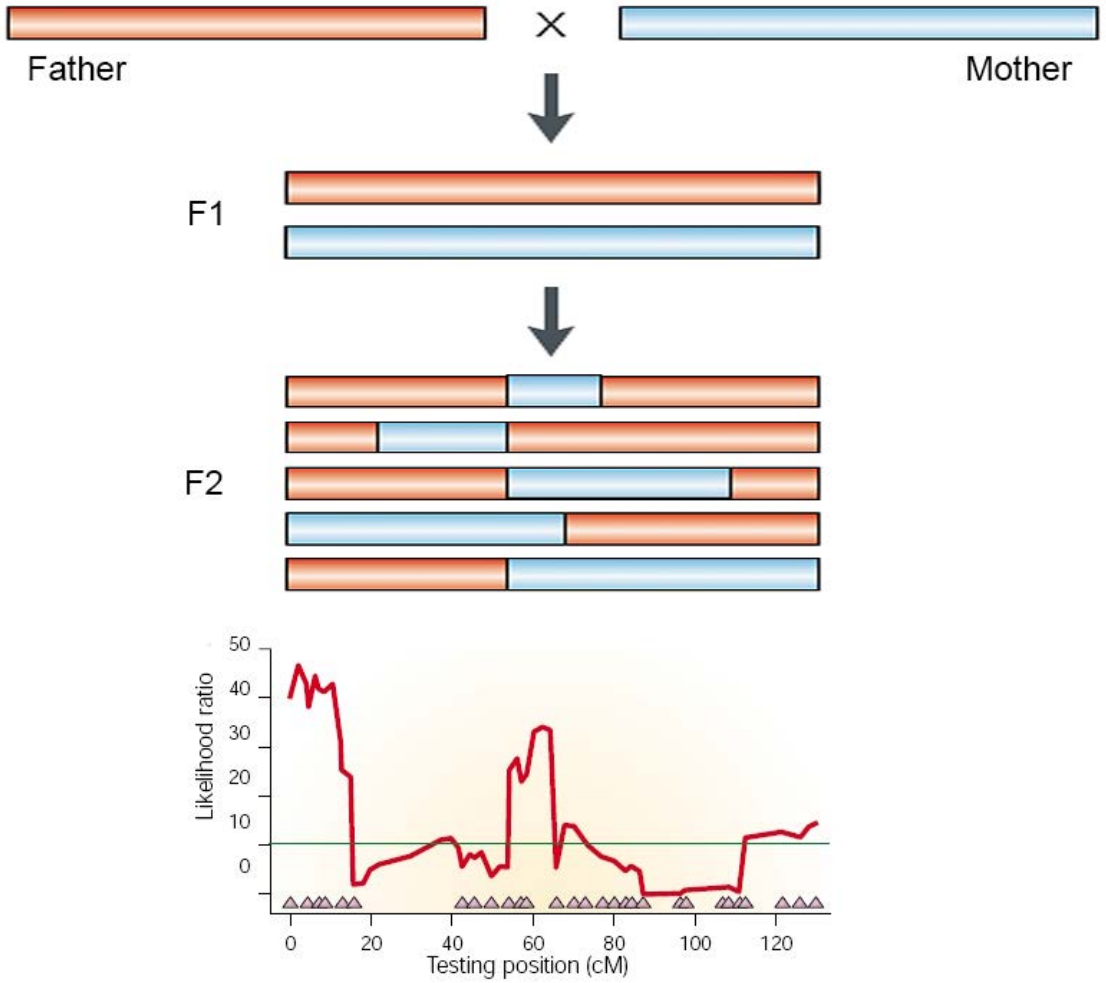


Figure 1.7. Quantitative trait locus mapping (adapted from Mackay, 2001). Two divergent populations for a specific trait are crossed to create a F1 population that contains the chromosomes of both progenitors (red and blue bars). F1 individuals are crossed among themselves and produce a F2 population. F2 individuals contain different fractions of the genome of each parental line. Phenotypic traits of each F2 individual are collected and then, associated with molecular markers that cover their entire genome (plot).

All the porcine QTLs published are archived in the pig QTL database (PigQTLdb) (Hu *et al.*, 2016). Of the more than 27,000 QTLs that are contained in the PigQTLdb, around 15,000 QTLs are related with meat and carcass traits (https://www.animalgenome.org/cgi-bin/QTLdb/SS/index; Release 36 accessed August 2018).

One of the first candidate genes found in pig was the ryanodine receptor 1 (*RYR1*), which is associated with leanness and susceptibility to stress (Fujii *et al.*, 1991). Since then, other candidate genes and even causal mutations have been found in association with pig meat quality traits through QTL mapping and GWAS analysis: the ELOVL fatty acid elongase (*ELOVL6*) gene with muscle and backfat percentages of palmitic and palmitoleic acids (Corominas *et al.*, 2013b); the genes of the fatty acid binding proteins 4 and 5 (*FABP4* and *FABP5*) with the FA deposition and growth (Estellé *et al.*, 2006; Ojeda *et al.*, 2006); the insulin like growth factor 2 (*IGF2*) gene with muscle growth, fat deposition and carcass weight (Van Laere *et al.*, 2003; Estellé *et al.*, 2005); the protein kinase AMP-activated non-catalytic subunit gamma 3 (*PRKAG3*) with glucose metabolism in muscle (Milan *et al.*, 2000), and the stearoyl-CoA desaturase (*SCD*) gene with MUFA content in pork (Estany *et al.*, 2014).

QTL mapping and GWAS analysis also possess some drawbacks. Certain populations have long-range linkage disequilibrium regions, which are genomic regions with a reduced variability due to selective sweep or artificial selection. These phenomena cause that SNPs that are relatively far from the QTL may present an association to the trait and thus, precision is reduced (Goddard and Hayes, 2009). In addition, certain associations or causal mutations may be specific of a population and thus, may not be replicated in another population.

1.7 Microbiomics

Microbiomics has potential applications in animal breeding. Microbiota influences methane emission (Su *et al.*, 2014), feed conversion ratio (Camarinha-Silva *et al.*, 2017) or disease resistance (Kim and Isaacson, 2017), among other characters. Statistical models that include microbiota abundances improve their prediction accuracies for complex traits (Camarinha-Silva *et al.*, 2017). In addition, the heritability of microbial genus abundances in the pig gut has been showed to range from low to high values (Estellé *et al.*, 2014; Camarinha-Silva *et al.*, 2017). Consequently, the microbiota of the parents is partially transmitted to their offspring and thus, it is susceptible of be selected and improved. However, while host genetics has been proposed as an important factor in the determination of the gut microbiota (Turpin *et al.*, 2016), it seems to have less impact than age, diet or environment (Spor *et al.*, 2011; Rothschild *et al.*, 2018).

The characterization of the entire microbiota of a biological sample is carried out through molecular methods such as the analysis of 16S rRNA genes, 18S rRNA genes, or other marker genes and genomic regions. Then, each amplified sequence is assigned to a microbial taxon (bacteria, archaea, or lower eukaryotes) at different taxonomic levels from phylum to species (Marchesi and Ravel, 2015).

The advancement of NGS technologies allowed the establishment in 2004 of the whole-metagenome shotgun sequencing method, which is used to sequence the genome of thousands of organisms in parallel (Tyson *et al.*, 2004; Venter *et al.*, 2004). Furthermore, one of the advantages of whole-metagenome shotgun sequencing over 16S rRNA sequencing is the detection of scarcer microorganisms and a better taxa assignation. Finally, through whole-metagenome shotgun sequencing, the first reference gene catalogue of the pig gut microbiome was obtained (Xiao *et al.*, 2016).

1.7.1 Microbiota profiling through the 16S rRNA gene sequencing

The most widely used method to assign taxonomies has been the analysis of the relatively well-conserved 16S rRNA genes in mixtures of organisms. Around

1,500 base pairs (bp) is the average length of the 16S rRNA gene which is transcribed into 16S ribosomal RNA that forms part of the 30S small subunit of a prokaryotic ribosome (Woese *et al.*, 1983). The 16S rRNA gene contains nine hypervariable regions (V1-V9, Figure 1.8) ranging from 30 to 100 bp long which are useful to differentiate taxa (Gray *et al.*, 1984). Primers are located inside conserved regions of the 16S rRNA gene and these hypervariable regions are sequenced to reconstruct phylogenies and assign taxonomies.

The 16S rRNA gene sequencing method was firstly used by Carl R. Woese in 1977 to define the three-domain system and classify all the cellular life forms into archaea, bacteria, and eukaryote domains (Woese and Fox, 1977; Woese, 1987). The first universal primers were devised in 1990 using consensus sequences for the 16S rRNA genes of different microorganisms (Giovannoni *et al.*, 1990; Ward *et al.*, 1990; Weisburg *et al.*, 1991). Since that time, the number of publications describing the composition and structure, and sometimes function, of the microbial communities increased exponentially.

In pigs, studies using the 16S rRNA gene sequencing technique have been published analysing the pig microbiome in different locations or biological states. For example, the microbiota of the nasal cavity (Slifierz *et al.*, 2015), the skin (McIntyre *et al.*, 2016) and the digestive tract (Zhao *et al.*, 2015; Yang *et al.*, 2016; Kelly *et al.*, 2017) have been described. Other works analysed the microbiota found in luminal contents of pigs with different ages (Zhao *et al.*, 2015) or extreme fatness (Yang *et al.*, 2016) and feed conversion ratios (Quan *et al.*, 2018).

One of the disadvantages of 16S rRNA gene sequencing is the need to use a reference to contrast the sequences obtained and perform the taxonomy assignation. In this scenario, taxonomic references for the 16S rRNA technology are better for human microbiota because reference databases are usually made from culture-based approaches using human samples. However, it is estimated that 20% to 60% of the human-associated microbiome is uncultivable, resulting in an underestimation of its diversity (NIH HMP Working Group *et al.*, 2009). Furthermore, taxonomic assignment is highly sensitive to the region and length of the 16S rRNA gene sequenced (Liu *et al.*, 2008; Guo *et al.*, 2013). Nevertheless, most regions provide stable estimates of phylum abundances (Liu

et al., 2008). In the next years, the lowering prices of sequencing will allow a cheaper and complete sequencing of the 16S rRNA gene that will increase the accuracy of taxonomic assignments.

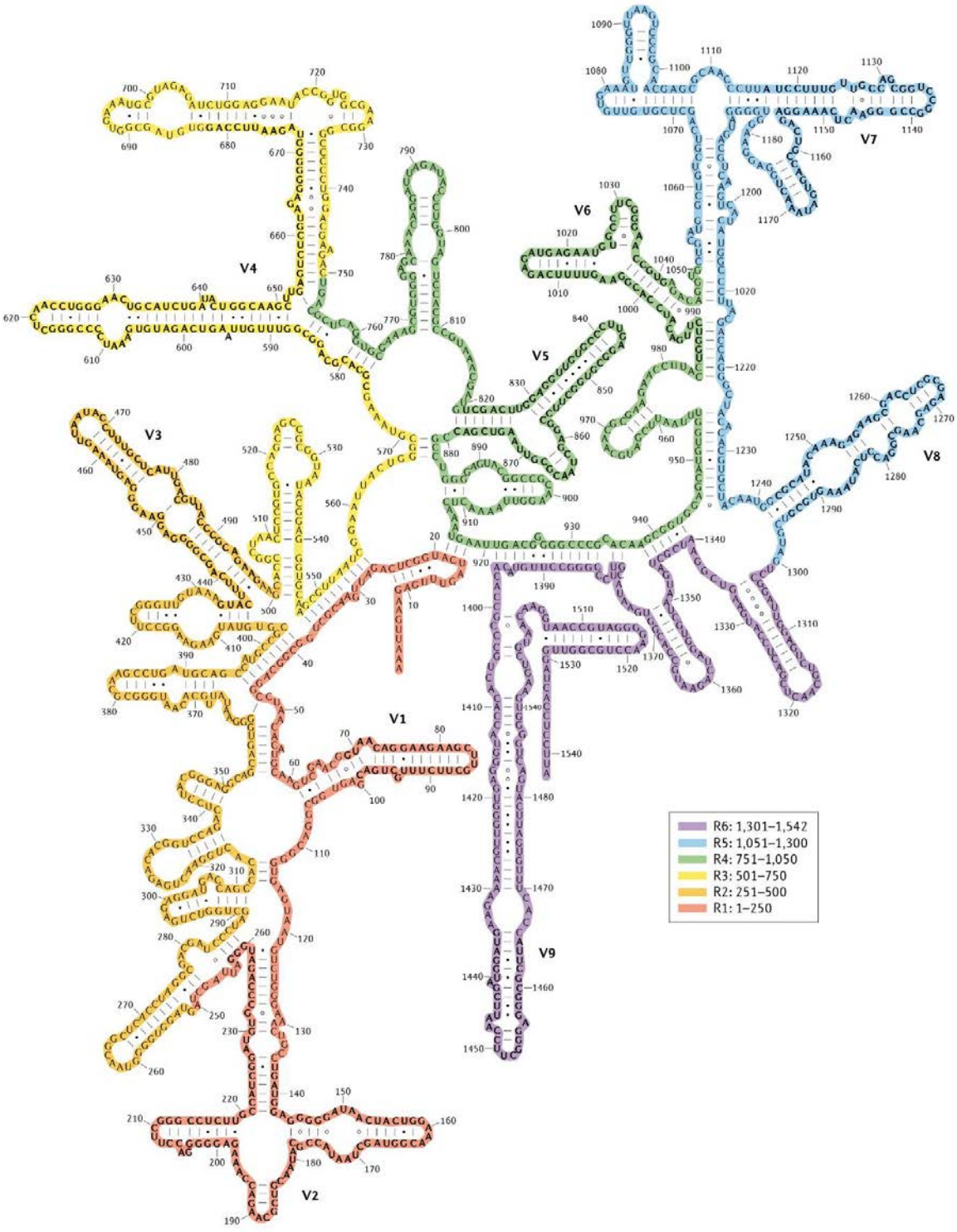


Figure 1.8. Variable regions of the 16S ribosomal RNA (Yarza *et al.*, 2014). In bold are marked the hypervariable regions from V1 to V9.

1.7.2 Anatomy and physiology of the pig gut

The gastrointestinal tract of pigs and humans share many anatomical, immunological and physiological similarities with some distinct differences (Gonzalez *et al.*, 2015; Wang and Donovan, 2015). Moreover, pigs and humans microbiomes share more non-redundant genes than human and mouse (Xiao *et al.*, 2016). In this regard, pigs are appropriate alternative models to study human digestive diseases. However, the dog and human microbiomes are even more similar in gene content and response to diet (Coelho *et al.*, 2018).

The small intestine of the pig starts in the pylorus and ends at the ileocecal valve. It is divided in three sections: duodenum, jejunum and ileum. The first section, the duodenum, is one meter long (Climent *et al.*, 2005) and is involved in the digestion of proteins, monosaccharides and SCFAs (Scheithauer *et al.*, 2016). The jejunum and the ileum measure 17 meters in total (Climent *et al.*, 2005), however its separation is not anatomically distinct (Young *et al.*, 2006). The term jejunum derives from the Latin "*jejunus*" which means "empty of food" as it is found empty at death (Climent *et al.*, 2005). The jejunum is responsible for the absorption of free FAs, calcium and fat-soluble vitamins (Scheithauer *et al.*, 2016). The last part of the small intestine, the ileum, absorbs bile salts, vitamin B_{12} and the rest of the digested products that were not absorbed previously (Scheithauer *et al.*, 2016).

The large intestine of pigs is five meters long on average ranging from four to six meters (Gonzalez *et al.*, 2015). The caecum measures 0.4 meters whereas the colon and the rectum measure 4.6 meters (Climent *et al.*, 2005). Apart from the absorption of water, the large intestine is involved in the fermentation of SCFAs (Scheithauer *et al.*, 2016). The reduced amount of simple nutrients and the lack of oxygen in the large intestine increase the number of microorganisms able to ferment non-digestible carbohydrates such as starch, plant cell wall and oligosaccharides (Cummings and Macfarlane, 1991; Louis *et al.*, 2007) (Figure 1.9).

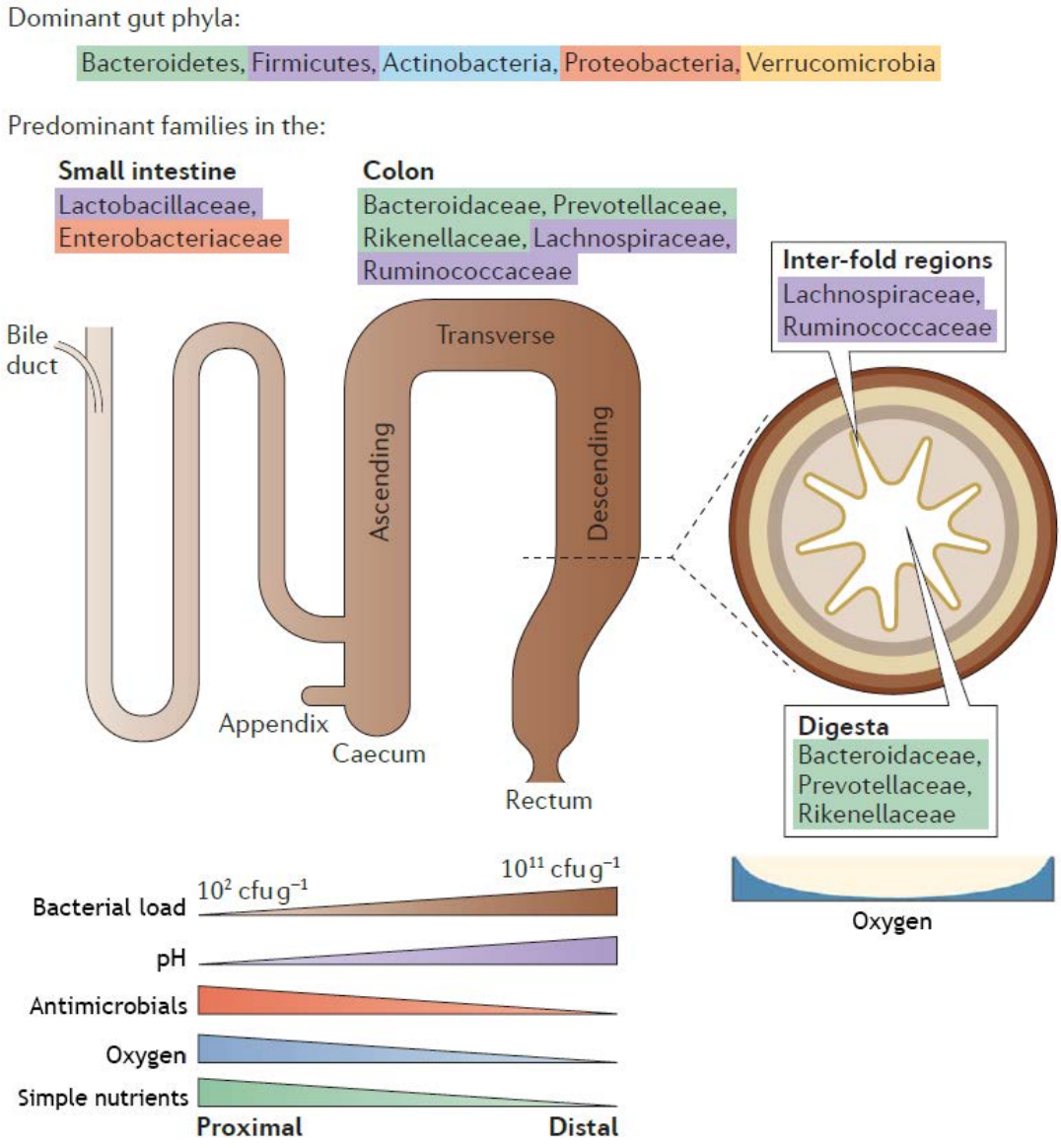


Figure 1.9. Spatial heterogeneity of the gut microbiota in the human lower gastrointestinal tract (adapted from Donaldson, Lee and Mazmanian, 2016, and Pereira and Berry, 2017).

1.8 European swine breeds

During the early Pliocene (5.3-3.5 million years ago), pigs (*Sus scrofa*) emerged in South East Asia (Groenen *et al.*, 2012). Around 10,000 years ago, pigs were independently domesticated in two different regions, western Eurasia and East Asia (Larson *et al.*, 2005). In addition, there was an exchange of genetic material between European and Asian pig breed populations and especially in Europe,

between wild and domesticated individuals (White, 2011; Groenen *et al.*, 2012). Therefore, since the pig domestication process started, selective breeding has led to the different swine breeds known today.

The most important and produced commercial swine breeds in Europe can be divided into three different groups considering their selective and crossbreeding schemes due to their productive aptitudes (Tibau, 1992; Sañudo Astiz, 2008):

The first group are formed by breeds characterized by their prolificacy and food conversion, fundamentally dedicated to act as maternal lines. This group is composed by Large White, Landrace and Yorkshire breeds that produce hybrid females and finisher males used to increase meat production.

The second group is formed by Belgian Landrace and Pietrain breeds that are used exclusively as paternal lines due to their excellent carcass conformation. These two breeds are characterized by a better production of lean meat and an efficient conversion rate (Kouba and Sellier, 2011). However, they are less prolific than the first group breeds.

The third group is composed by Duroc and Hampshire breeds that are used to improve carcass meat quality as paternal lines due to their high percentage of intramuscular fat (Kouba and Sellier, 2011). Furthermore, they are characterized by their rusticity, with a good conformation and low food consumption. However, they present lower growth rates and tend to fatten up. Therefore, these breeds are usually crossed with maternal line breeds such as Large White or Landrace.

Other relevant breed produced in Spain is the autochthonous Iberian pig. This is a rustic breed, dark or red-haired coated, with a less productive conformation than commercial breeds. In addition, Iberian pigs develop a higher proportion of intermuscular fat than the commercial breeds, which contain less PUFAs than the intramuscular fat of Duroc pigs (Benítez *et al.*, 2017). Iberian pig products such as ham possess an excellent organoleptic quality due to the high fat infiltration rate in Iberian pig muscle, with an increased quantity of MUFAs (mainly oleic acid) and a reduced proportion of PUFAs (Serra *et al.*, 1998).

1.9 The IBMAP Consortium

In 1996, the IBMAP consortium was formed as a cooperative effort by the *Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria* (INIA), the *Institut de Recerca i Tecnologia Agroalimentàries* (IRTA) and the *Universitat Autònoma de Barcelona* (UAB). The objective of this consortium was the identification of QTLs associated with pork quality and growth. Due to its excellent meat quality (Serra *et al.*, 1998), boars of the Iberian pig breed were used as male parental line and crossed with sows of other three breeds (Landrace, Duroc and Pietrain). The F1 obtained was backcrossed again with sows of their respective maternal line (Figure 1.10). In addition, other F2 and F3 crosses were also performed.

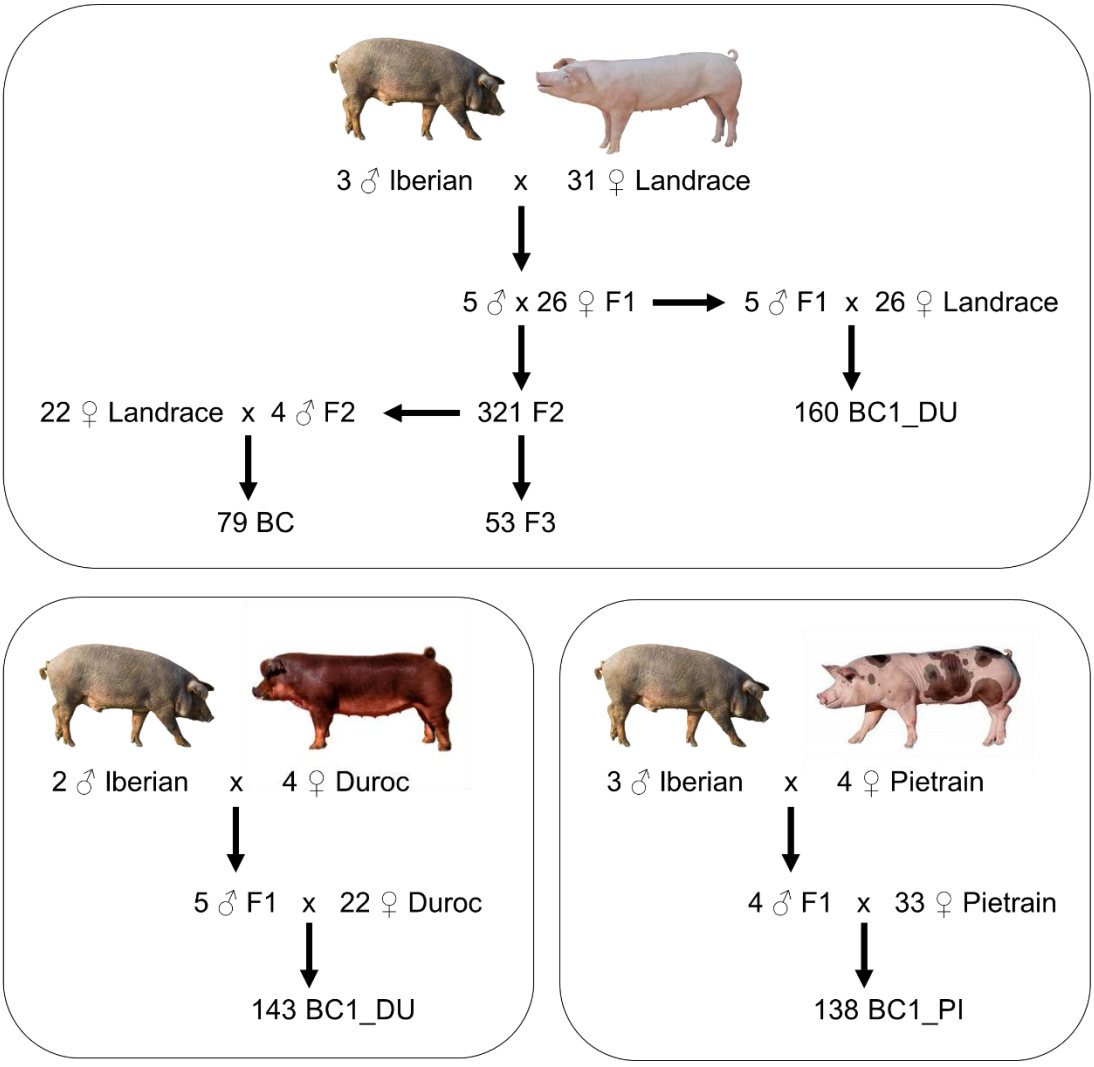


Figure 1.10. Schematic representation of the three IBMAP backcrosses (Landrace, Duroc and Pietran).

The first studies of the IBMAP group identified several QTLs associated with carcass quality, growth, fatness and FA composition on *Sus scrofa* chromosomes (SSC) SSC2, SSC3, SSC4, SSC6, SSC7, SSC8, SSC10, SSC12 and SSCX, using microsatellites as molecular markers (Ovilo *et al.*, 2000, 2002, 2005, Pérez-Enciso *et al.*, 2000, 2002, 2005; Varona *et al.*, 2002; Clop *et al.*, 2003; Mercadé *et al.*, 2005, 2006; Muñoz *et al.*, 2007). A QTL genome scan for the backfat FA composition in pigs was performed in a F_2 Iberian × Landrace cross using microsatellites, detecting significant QTLs on SSC4 for linoleic acid (C18:2(n-6)) and double-bond and peroxidability indices; on SSC6 for double-bond and unsaturated indices; on SSC8 for palmitic acid (C16:0), palmitoleic acid (C16:1(n-7)) and average chain length; on SSC10 for myristic acid (C14:0), and on SSC12 for linolenic acid (C18:3(n-3)) (Clop *et al.*, 2003).

With the new high-throughput genotyping platform from *Illumina* (PorcineSNP60 BeadChip), the IBMAP group improved the resolution of the previous QTLs described and found new genomic regions associated with the previous traits (Fernández *et al.*, 2012; Ramayo-Caldas *et al.*, 2012a; Corominas *et al.*, 2013b; Muñoz *et al.*, 2013; Revilla *et al.*, 2014). Ramayo-Caldas *et al.* (2012a) identified a total of 43 pig chromosomal regions associated with FA composition traits in IMF of 144 BC1_LD pigs using GWAS, and *ELOVL7* (ELOVL fatty acid elongase 7) was proposed as a candidate gene to explain the variation on the arachidic acid (C20:0) content and the gondoic acid/arachidic acid (C20:1(n-9)/C20:0) ratio on SSC16. In addition, the association between a genomic region on SSC8 and the percentages of palmitic acid (C16:0), palmitoleic acid (C16:1(n-7)), SFAs, and the C16:1(n-7)/C16:0 and C18:1(n-7)/C16:1(n-7) ratios was also found (Ramayo-Caldas *et al.*, 2012a). This genomic region on SSC8 was later refined by Corominas *et al.* (2013b), using a combination of GWAS and LDLA (linkage disequilibrium and linkage analysis), which reduced the confidence interval of the SSC8 QTL for the intramuscular C16:0 and C16:1(n-7) content in the BC1_LD pigs. In the same study, a polymorphism in the *ELOVL6* gene (*ELOVL6:c.-533C>T*) was highly associated with the percentages of C16:0 and C16:1(n-7) in backfat and IMF of BC1_LD animals. Later on, *ELOVL6* was confirmed as the candidate gene on SSC8 to explain the variation on the percentages of C16:0 and C16:1(n-7) in backfat and IMF in the BC1_LD animals, detected by linkage

and GWAS approaches (Muñoz *et al.*, 2013; Revilla *et al.*, 2014). In a later study, the *ELOVL6:c.-394G>A* SNP was proposed as the causal mutation to explain the QTL on SSC8 that affects FA composition (Corominas *et al.*, 2015).

The development of NGS technologies allowed the sequencing of whole genomes and transcriptomes in the IBMAP population. RNA-Seq was used to identify differentially expressed genes in the principal tissues involved in lipid metabolism: liver (Ramayo-Caldas *et al.*, 2012b), adipose tissue (backfat) (Corominas *et al.*, 2013a) and muscle (*Longissimus dorsi*) (Puig-Oliveras *et al.*, 2014). To this end, BC1_LD female pigs were divided in two extreme groups depending on their intramuscular FA profile, one group had higher SFA and MUFA content and the other one had higher PUFA content. The *SCD* gene was differentially expressed in muscle and backfat, while other genes such as *ELOVL6* and *FASN* were only differentially expressed in backfat.

In addition, the expression of a set of potential candidate genes involved in lipid metabolism was studied in muscle (Puig-Oliveras *et al.*, 2016), liver (Ballester *et al.*, 2017) and backfat (Revilla *et al.*, 2018), using quantitative reverse transcription PCR (RT-qPCR) in the BC1_LD pigs. In backfat, significant expression QTLs (eQTLs) for the differences in the expression of *ELOVL6* and *FADS2* (fatty acid desaturase 2) genes were found, including one *FADS2* eQTL on SSC2 where the *FADS2* gene is located, indicating that a polymorphism in the *FADS2* gene may be regulating its expression (Revilla *et al.*, 2018).

Finally, in a recent study, 1,279 CNVs were found using WGS data from two Iberian boars and five Landrace sows, founders of the BC1_LD backcross (Revilla *et al.*, 2017). This WGS data also has been useful to find new polymorphisms in the BC1_LD pigs through SNP calling.

2. OBJECTIVES

This PhD thesis was done under the framework of the AGL2014-56369-C2-2-R and AGL2017-82641-R projects funded by the *Ministerio de Economía y Competitividad* (MINECO).

The main goal of this thesis was to increase the knowledge of the genetic basis determining the fatty acid profile and the gut microbiota composition in pigs.

The specific objectives were:

1. To identify chromosomal regions and positional candidate genes associated with the backfat and intramuscular fatty acid composition in pigs with three different genetic backgrounds but having the Iberian breed in common.

2. To detect indels from Whole Genome Sequence data of founders of the IBMAP population, a cross between Iberian and Landrace breeds, and to study its association with intramuscular fatty acid composition.

3. To determine the composition and interactions of the microbiota along different sections of the Iberian pig gut and to evaluate their metagenome functions.

4. To describe host-microbiota interactions in pigs and to find genomic regions associated with gut microbiota composition.

3. ARTICLES AND STUDIES

Paper I

Genome-Wide Association Study for backfat and intramuscular fatty acid composition in three different pig crosses based on the Iberian breed

Crespo-Piazuelo, D., Lourdes Criado-Mesas, L., Revilla, R., Castelló, A., Noguera, J. L., Fernández, A. I., Ballester, M., and Folch, J. M.

Genetics Selection Evolution (*submitted*)

Genome-Wide Association Study for backfat and intramuscular fatty acid composition in three different pig crosses based on the Iberian breed

Daniel Crespo-Piazuelo[1,2], Lourdes Criado-Mesas[2], Manuel Revilla[1,2], Anna Castelló[1,2], José L. Noguera[3], Ana I. Fernández[4], Maria Ballester[5] and Josep M. Folch[1,2]

[1] Plant and Animal Genomics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, 08193 Bellaterra, Spain

[2] Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Spain

[3] Genètica i Millora Animal, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), 25198 Lleida, Spain

[4] Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 28040 Madrid, Spain

[5] Departament de Genètica i Millora Animal, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), 08140 Caldes de Montbui, Spain

Address for correspondence:

D. Crespo-Piazuelo, Plant and Animal Genomics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, Edifici CRAG, Campus UAB, 08193 Bellaterra, Spain

E-mail: daniel.crespo@cragenomica.es

Abstract

Background

Meat quality can be modified by the fatty acid (FA) composition and the amount of fat that is contained in adipose tissue and muscle. The variability of FA composition traits has an important genetic component and shows moderate to high heritability values. The objective of the present study was to find genomic regions associated with the FA composition in adipose tissue (backfat) and muscle using genome-wide association studies in pigs with three different genetic backgrounds but having the Iberian breed in common. The Iberian breed is known by its great meat quality, characterised by a high intramuscular fat content and an increased proportion of monounsaturated FAs (mostly oleic fatty acid) and a reduced quantity of polyunsaturated FAs.

Results

The genotypes of 38,424 common SNPs covering the pig genome were obtained in 441 pigs with different genetic backgrounds. Genome-wide association studies were performed among these SNPs and 60 phenotypic traits related to FA composition in backfat and muscle (*Longissimus dorsi*). Nine significant associated regions were found in backfat in the *Sus scrofa* chromosomes (SSC): SSC1, SSC2, SSC4, SSC6, SSC8, SSC10, SSC12, and SSC16. For the intramuscular fat, six significant associated regions were identified in SSC4, SSC13, SSC14, and SSC17. A total of 50 candidate genes were proposed to explain the variation in backfat and muscle FA composition traits, including six genes (*ELOVL3, ELOVL6*, *ELOVL7*, *FADS2*, *FASN* and *SCD*) which have been previously described in GWAS performed on different pig populations.

Conclusion

Our study increases the knowledge of the genetic basis of FA composition and lipid metabolism. Molecular markers in candidate genes found in our study may be useful for the selection for meat quality of commercial pig breeds and the study of human diseases. Several polymorphisms were breed-specific, and further analysis would be needed to evaluate possible causal mutations.

Keywords

adipose tissue; fatty acid composition; GWAS; lipid metabolism; muscle; pig; QTLs

Background

Meat quality depends on the consumer's perception, which is subjected to the socio-demographic backgrounds of the consumer [1], and is based on factors such as the nutritional value and the organoleptic properties of meat [2]. These factors can be modified by the fatty acid (FA) composition and the amount of fat that is contained in adipose tissue and muscle [3]. In addition, the consumer is becoming more concerned about the healthfulness of meat [1]. Certain saturated FAs (SFAs) raise the cholesterol and low-density lipoprotein (LDL) levels in blood, which increase the risk of suffering a cardiovascular disease [4, 5]; whereas monounsaturated FAs (MUFAs) reduce plasma total LDL-cholesterol without affecting high-density lipoprotein (HDL) levels, which have an anti-atherogenic effect [4, 5]. Polyunsaturated FAs (PUFAs), especially long-chain omega-3 fatty acids, also reduce LDL-cholesterol levels and the risk of cardiovascular disease [6], but high amount of PUFAs in meat increase its susceptibility to oxidation, producing meat with undesirable sensory properties [3]. In contrast, a high concentration of MUFAs improve meat flavour [7].

The Iberian pig is a breed characterized by its great meat quality, due to its high intramuscular fat (IMF) content with an increased proportion of MUFAs (mostly oleic acid) and a reduced quantity of PUFAs [8, 9]. Conversely, other commercial breeds such as Pietrain and Landrace produce lean meat with a lower proportion of fat, although Pietrain carcasses exhibit a high ratio of intermuscular to subcutaneous fat [10]. Duroc pigs also exhibit a high intermuscular fat weight and develop a higher proportion of intramuscular fat than the commercial breeds [10], which contain more PUFAs than the Iberian breed [11]. Altogether, the FA composition of adipose tissue and muscle in pigs show moderate to high heritability values [7, 9, 12], revealing the importance of the genetic component in the variability of FA composition traits. In addition, FA composition is an

expensive trait to measure and often require the slaughter of the animals. Polymorphisms associated with these traits can be used as genetic markers to evaluate the breeding value of an animal and increase the rate of genetic gain [13].

Genetic selection in pigs has been intensifying thanks to high-density genotyping platforms, such as the PorcineSNP60 BeadChip (*Illumina*®) [14] or the Axiom™ Porcine Genotyping Array (Affymetrix, Inc.) [15]. These chips allow the genotyping of markers distributed along the pig genome to perform Genome-Wide Association Studies (GWAS) for production traits. Through the use of GWAS, significant Quantitative Trait Loci (QTLs) in the pig genome have been identified for the FA composition in adipose tissue and muscle in several populations of crossed and purebred pigs such as Duroc, Landrace, Large White, and Erhualian [16, 17, 26, 18–25]. In these studies, strong candidate genes related with lipid metabolism have been found for the FA composition in adipose tissue and muscle: ELOVL fatty acid elongases 6 and 7 (*ELOVL6* [17, 18, 21, 22, 24, 25] and *ELOVL7* [16, 19, 21, 22]), fatty acid synthase (*FASN*) [21–23], and stearoyl-CoA desaturase (*SCD*) [19–26].

In this context, our group generated different crosses among Iberian purebred pigs and commercial breeds (IBMAP population) to identify QTLs associated with the FA composition in adipose tissue and muscle. In previous works, pigs from an IBMAP experimental backcross ((Iberian x Landrace) x Landrace) were genotyped with the PorcineSNP60 BeadChip (*Illumina*®) array [14] and GWAS were performed for FA composition traits in adipose tissue and muscle using the Sscrofa10.2 assembly [16, 18]. In addition, GWAS have also been used in the IBMAP population to find QTLs associated with the expression of genes involved in lipid metabolism in adipose tissue and muscle [27, 28].

The present study aimed to identify genomic regions associated with FA composition of backfat and muscle in pigs with three different genetic backgrounds but having the Iberian breed in common using GWAS analyses.

Material and Methods

Ethics statement

All animal procedures were performed according to the Spanish Policy for Animal Protection RD1201/05, which meets the European Union Directive 86/609 about the protection of animals used in experimentation. The experimental protocol was approved by the Ethical Committee of IRTA (*Institut de Recerca i Tecnologia Agroalimentàries).*

Animal Material

All animals used in the present work belong to three distinct pig backcrosses: (Iberian x Landrace) x Landrace (BC1_LD, n=160), (Iberian x Duroc) x Duroc (BC1_DU, n=143), and (Iberian x Pietrain) x Pietrain (BC1_PI, n=138). Pigs were raised in an intensive system and fed *ad libitum* with a cereal-based commercial diet until slaughtered at 187.4±10.1 days of age on NOVA GENÈTICA S. A. experimental farm (Lleida, Spain). Detailed information of generation schemes, diet, growth, and housing conditions of the three backcrosses is described in Martínez-Montes *et al*. [29].

Samples of adipose tissue (backfat), diaphragm and *longissimus dorsi* muscle were collected at the commercial abattoir, snap-frozen in liquid nitrogen and stored at -80ºC. A gas chromatography of methyl esters protocol [30] was used to measure the FA profile (15 backfat FAs and 17 IMF FAs) of backfat samples taken between the third and the fourth ribs and 200 g of *longissimus dorsi* muscle. Then, the percentage of each individual FA was calculated, as well as the total percentage of SFA, MUFA, and PUFA.

In total, 60 traits were analysed in backfat and IMF: 32 traits for FA percentages and 28 traits for indices of FA metabolism, including FA ratios for the activities of desaturases and elongases (Table 1).

Table 1. Descriptive statistics including mean and SD of intramuscular fat and backfat fatty acid composition and FA indices in the merged dataset of the three backcrosses.

Group	Trait	Name	Backfat		Intramuscular fat	
			Mean	SD	Mean	SD
SFA	C14:0	Myristic acid	1.14	0.11	1.17	0.24
	C16:0	Palmitic acid	23.06	1.48	22.99	1.51
	C17:0	Margaric acid	0.37	0.10	0.27	0.10
	C18:0	Stearic acid	14.76	1.79	14.21	1.44
	C20:0	Arachidic acid	0.28	0.05	0.26	0.11
MUFA	C16:1(n-9)	7-Hexadecenoic acid	0.38	0.10	0.36	0.11
	C16:1(n-7)	Palmitoleic acid	1.53	0.25	2.61	0.50
	C17:1	Heptadecenoic acid	0.25	0.06	0.23	0.10
	C18:1(n-9)	Oleic acid	39.39	2.67	37.08	5.78
	C18:1(n-7)	Vaccenic acid	1.00	0.67	3.91	0.33
PUFA	C20:1(n-9)	Gondoic acid	1.02	0.16	0.82	0.20
	C18:2(n-6)	Linoleic acid	14.76	2.78	11.92	5.01
	C18:3(n-3)	α-Linolenic acid	0.77	0.13	0.50	0.23
	C20:2(n-6)	Eicosadienoic acid	-	-	0.43	0.28
	C20:3(n-3)	Eicosatrienoic acid	-	-	2.53	2.04
	C20:3(n-6)	Dihomo-γ-linolenic acid	0.76	0.12	0.51	0.14
	C20:4(n-6)	Arachidonic acid	0.16	0.05	0.22	0.13
Metabolic Ratios	SFA	Saturated fatty acids	39.60	2.86	38.90	2.44
	MUFA	Monounsaturated fatty acids	43.57	2.39	44.78	6.25
	PUFA	Polyunsaturated fatty acids	15.69	2.93	15.88	7.26
	MUFA/SFA	Ratio of MUFA to SFA	1.11	0.11	1.15	0.16
	PUFA/SFA	Ratio of PUFA to SFA	0.40	0.10	0.42	0.22
	MUFA/PUFA	Ratio of MUFA to PUFA	2.88	0.56	3.47	1.60
FA ratios	C16:1(n-7)/C16:0		0.07	0.01	0.11	0.02
	C18:1(n-7)/C16:1(n-7)		0.67	0.45	1.55	0.32
	C18:1(n-9)/C18:0		2.71	0.40	2.63	0.44
	C18:2(n-6)/C18:3(n-3)		19.29	1.99	26.27	11.76
	C20:1(n-9)/C20:0		3.72	0.46	3.42	1.09
	C20:3(n-6)/C18:2(n-6)		0.05	0.01	0.03	0.01
	C20:4(n-6)/C18:2(n-6)		0.01	0.00	0.19	0.08
	C20:4(n-6)/C20:3(n-6)		0.20	0.06	5.62	2.16

DNA extraction and SNP genotyping

DNA extraction was carried out with the phenol-chloroform method [31] from the diaphragm of the 441 pigs. DNA concentration and quality were measured with a Nanodrop® Spectrophotometer (ND-1000).

Two distinct platforms were used for genotyping. The PorcineSNP60 BeadChip (*Illumina*®) array [14] was employed to genotype 64,232 SNPs in the BC1_LD and BC1_PI animals using the Infinium® HD Assay Ultra protocol (*Illumina*®) and were visualized through the GenomeStudio software (2011.1 version, *Illumina*®). The Axiom™ Porcine Genotyping Array (Affymetrix, Inc.) [15] was used for genotyping 658,692 SNPs in the BC1_DU pigs and were visualized and filtered with the Axiom™ Analysis Suite 2.0. For the GWAS analysis, we only considered the 45,845 SNPs that were found in common between both platforms and mapped in the Sscrofa11.1 assembly. SNPs with a minor allele frequency (MAF) < 5% or with missing genotypes > 5% were removed using the PLINK software [32] (1.90b5 version). Finally, a total of 38,424 SNPs remained for further analysis.

Furthermore, thirteen SNPs in positional candidate genes were genotyped in the 441 pigs using Taqman OpenArray™ genotyping plates custom-designed in a QuantStudio™ 12K flex Real-Time PCR System (ThermoFisher Scientific). Of these thirteen SNPs, seven were *ELOVL6*-SNPs (rs325491325, rs345025813, rs341847499, rs322956047, *ELOVL6:c.416C>T*, *ELOVL6:c.1408C>T* and *ELOVL6:c.1922C>T*) [17, 33], one was a SNP (rs335191239) of the phospholipase A2 group XIIA (*PLA2G12A*) gene, three were *ELOVL7*-SNPs (rs322657523, rs343494956 and rs325490947), and two were SNPs (rs331708297 and rs322671019) located on the phosphoinositide-3-kinase regulatory subunit 1 (*PIK3R1*) gene.

Genome-Wide Association Studies (GWAS)

For the 441 pigs of the three backcrosses, GWAS were carried out between the 38,424 filtered SNPs and the fatty acid composition and metabolic indices in backfat and IMF, described in Table 1. Thus, inside an Unix environment, the GEMMA software [34] (0.96 version) was used to perform an univariate linear mixed model following this formula:

$y_{ijklm} = Sex_i + Batch_j + Backcross_k + \beta c_l + u_l + \lambda a_m + e_{ijklm}$,

where y_{ijklm} indicates the vector of phenotypic observations in the l^{th} individual; sex (two categories), batch (14 categories) and backcross (3 categories) are fixed effects; β is a covariate coefficient with c being carcass weight; u_l is the infinitesimal genetic effect considered as random and distributed as $N(0, A\sigma_u)$, where A is the numerator of the kinship matrix; λ_l is a -1, 0, +1 indicator variable depending on the l^{th} individual genotype for the m^{th} SNP; a_m represents the additive effect associated with the m^{th} SNP; and e_{ijklm} is the residual effect.

The same mixed model was used for the association analysis with thirteen SNPs located in positional candidate genes. Eight candidate gene-SNPs of the *Sus scrofa* chromosome (SSC) 8 region were included in the GWAS for the percentages of C14:0, C16:0, and C16:1(n-7) in backfat. Five candidate gene-SNPs of the SSC16 region were included in the GWAS for the C20:1(n-9)/C20:0 ratio in backfat.

GWAS were also performed individually for each backcross following the previously described model without the backcross effect.

The false discovery rate (FDR) method of multiple testing described by Benjamini and Hochberg [35] was used to measure the statistical significance for association studies at genome-wide level. The significant association threshold was set at FDR≤0.1. If the distance between two significant SNPs was less than 10 Mb, they were grouped inside the same interval. In addition, intervals for different traits were merged if they overlapped.

The web based tool PhenoGram [36] was used to visualize the differences obtained in the GWAS results between the merged dataset and each individual backcross.

Gene annotation and consequence prediction

Associated regions were annotated at 1 Mb on each side of the previously defined intervals in the pig genome. The extraction of the genes contained in the associated regions was performed with the BioMart tool [37] from the Ensembl project (www.ensembl.org; release 92) using the Sscrofa11.1 reference assembly. Furthermore, functional predictions of the significant SNPs were

performed with the Variant Effect Predictor tool [35] from the Ensembl project (release 92).

Results and Discussion

GWAS results

A GWAS was performed using a total of 38,424 SNPs and the 60 phenotypic traits related with backfat and IMF FA composition (percentages, indices, and ratios) in a total of 441 pigs from three different backgrounds (BC1_DU, BC1_LD and BC1_PI). In backfat, 96 significant associated SNPs located in nine *Sus scrofa* chromosomal regions of SSC1, SSC2, SSC4, SSC6, SSC8, SSC10, SSC12, and SSC16 were significantly associated with 12 traits: myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1(n-7)), 7-hexadecenoic acid (C16:1(n-9)), oleic acid (C18:1(n-9)), linoleic acid (C18:2(n-6)), dihomo-γ-linolenic acid (C20:3(n-6)), MUFA, PUFA, MUFA/PUFA, PUFA/SFA, and the gondoic acid/arachidic acid (C20:1(n-9)/C20:0) ratio (FDR≤0.1; Table 2). In IMF, 39 SNPs located in six regions of SSC4, SSC13, SSC14 and SSC17 were significantly associated with 6 traits: stearic acid (C18:0), arachidic acid (C20:0), eicosatrienoic acid (C20:3(n-3)), MUFA/SFA, the oleic acid/stearic acid (C18:1(n-9)/C18:0) ratio, and the arachidonic acid/dihomo-γ-linolenic acid (C20:4(n-6)/C20:3(n-6)) ratio (FDR≤0.1; Table 3). The QTL located on SSC2 in backfat was the region associated with the highest number of traits (seven traits), whereas the rest of the regions in backfat and in IMF were associated with one or two traits, except for the QTL on SSC8 in backfat (three traits). The regions with the highest number of significant SNPs were the QTL on SSC16 in backfat (34 SNPs) and the first QTL on SSC14 (21 SNPs). The chromosome with the highest number of QTLs found was SSC4 with one QTL for the FA composition in backfat and two QTLs for the FA composition in IMF. In addition, no QTLs in common were found between backfat and IMF.

Table 2. Description of the regions associated with the fatty acid composition in backfat and the candidate genes contained within them.

Region	Chr	Start (bp)	End (bp)	No.SNPs	Top SNP	MAF	*p*-value	FDR	Trait	Candidate gene
BF1	1	145,957,212	147,979,672	2	rs80899816	0.172	1.38E-06	2.65E-02	C16:1(n-9)	*GALR1*
BF2	2	0	10,772,447	22	rs81306755	0.391	3.07E-09	9.77E-05	C16:1(n-9); C18:1(n-9); C18:2(n-6); MUFA; MUFA/PUFA*; PUFA; PUFA/SFA	*CPT1A*; *ESRRA*; *FADS1*; *FADS2*; *FADS3*; *IGF2*; *INS*; *PLA2G16*; *PNPLA2*; *SIRT3*; *ssc-mir-192*
BF3	4	80,446,371	82,565,259	4	rs80848071	0.415	6.01E-06	1.54E-02	MUFA/PUFA	*PRRX1*; *SLC19A2*
BF4	6	15,339,713	27,053,724	3	rs81322046	0.203	3.46E-07	1.33E-02	C20:3(n-6)	*BBS2*; *COQ9*; *GOT2*; *SLC12A3*
BF5	8	108,399,818	116,409,757	14	rs81403349	0.211	3.05E-10	1.17E-05	C14:0; C16:0*; C16:1(n-7)	*ELOVL6*; *HADH*; *PLA2G12A*
BF6	10	29,495,950	31,756,975	3	rs81423282; rs81423288	0.256	1.46E-05	6.24E-02	C16:1(n-9)	*NTRK2*; *RMI1*
BF7	10	51,874,658	53,920,353	2	rs80979357	0.095	7.46E-06	4.78E-02	C20:3(n-6)	*DNAJC1*; *PIP4K2A*
BF8	12	0	1,910,198	12	rs81308244	0.237	1.98E-10	7.59E-06	C14:0	*ASPSCR1*; *FASN*; *METRNL*; *NOTUM*
BF9	16	29,669,240	48,628,216	34	rs81297480; rs81458871	0.127	3.20E-08	6.14E-04	C20:1(n-9)/C20:0	*ELOVL7*; *GZMA*; *PIK3R1*; *PLPP1*

*Indicates the most significant trait. The *p*-value and FDR refer to this trait.

Table 3. Description of the regions associated with the fatty acid composition in intramuscular fat and the candidate genes contained within them

Region	Chr	Start (bp)	End (bp)	No.SNPs	Top SNP	MAF	*p*-value	FDR	Trait	Candidate gene
LD1	4	19,019,566	21,057,452	2	rs80910044	0.177	2.36E-06	8.53E-02	C20:4(n-6)/C20:3(n-6)	*ENPP2*; *EXT1*; *NOV*
LD2	4	122,756,546	124,979,309	3	rs81347340; rs80915252	0.437	1.18E-05	7.65E-02	C20:3(n-3)	*ABCD3*; *GCLM*
LD3	13	175,539,436	181,652,057	3	rs81441592	0.066	1.81E-06	6.96E-02	C20:3(n-3)	*LIPI*; *NRIP1*; *ssc-let-7c*
LD4	14	109,946,218	114,621,937	21	rs335655209	0.115	1.08E-07	7.79E-04	C18:1(n-9)/C18:0*; MUFA/SFA	*ELOVL3*; *SCD*
LD5	14	140,151,934	142,374,711	6	rs318740977; rs80814938; rs80883500	0.083	3.56E-06	3.26E-02	C18:0	*BNIP3*; *CYP2E1*; *ECHS1*
LD6	17	30,061,857	32,867,849	4	rs324135473	0.148	2.63E-06	9.09E-02	C20:0	*ABHD12*; *ACSS1*; *PANK2*; *ssc-mir-103-2*

*Indicates the most significant trait. The *p*-value and FDR refer to this trait.

In the following sections, the candidate genes mapped in the genomic regions associated with the phenotypic traits of backfat and IMF are discussed in detail. The list of candidate genes is summarized in Table 2 for backfat traits and in Table 3 for IMF traits. All the significantly associated SNPs, and their predicted consequences, for the FA composition in backfat and in IMF are listed on the Supplementary Tables S1 and S2, respectively.

QTLs for fatty acid composition in backfat and candidate genes

The 145.96-147.98 Mb region of SSC1 was associated with the C16:1(n-9) content in backfat (Table 2). The rs80899816 SNP was the most significant (*p*-value=1.38×10^{-6}) of the two SNPs comprised in this region, but both were located inside an intergenic region. The galanin receptor 1 (*GALR1*) gene was located in this region. This gene is a member of the galanin receptor family, which bind the neuropeptide hormone galanin [38]. Galanin regulates a range of biological functions such as food intake, neurogenesis, memory, and gut secretion [39]. Remarkably, the galanin-mediated signalling cascade has been associated with an activation of adipogenesis in high-fat diet induced obese mice [39, 40]. In zebrafish, *GALR1* was up-regulated if animals were fed with a high fat fodder or with linoleic acid, participating in the accumulation of lipid droplets in cells [41].

In SSC2, the 0-10.77 Mb region was associated with the abundance of three FAs in backfat (Table 2), C16:1(n-9), C18:1(n-9), and C18:2(n-6), and four metabolic ratios, MUFA, PUFA, MUFA/PUFA, and PUFA/SFA. In this QTL, 22 SNPs were found significantly associated with these seven traits and the most significant SNP was rs81306755 (*p*-value=3.07×10^{-9}) for the MUFA/PUFA ratio. This significant SNP was located inside an intron of a novel gene (*ENSSSCG00000014565*) that was orthologous of the IFITM (interferon-induced transmembrane) protein family. Fatty acid desaturases 1-3 (*FADS1*, *FADS2*, *FADS3*) are the most promising candidate genes found in this region to explain the variations in MUFA and PUFA content in backfat, specially *FADS2*. The essential FAs C18:2(n-6) and C18:3(n-3) are desaturated by FADS2, which can also desaturate C16:0 and C18:1(n-9) [42, 43]. We have previously described the association of this SSC2 with the percentages of C16:1(n-9), C18:2(n-6), C18:3(n-3), and PUFAS in backfat in the BC1_LD animals [44]. Furthermore,

FADS2 was proposed as a candidate gene to explain the variation of two metabolic indices (C20:3(n-6)/C18:2(n-6) and C20:4(n-6)/C20:3(n-6)) in the IMF of Erhualian pigs [22]. Other candidate gene, the *ssc-mir-192* gene, codifies for the miR-192 microRNA which impairs adipocyte triglyceride storage and suppress the production of another desaturase, SCD [45]. In addition, three candidate genes found in this region are related with the oxidation of different FAs, carnitine palmitoyltransferase 1A (*CPT1A*), estrogen related receptor alpha (*ESRRA*), and phospholipase A2 group XVI (*PLA2G16*). The *CPT1A* gene encodes for a protein that is involved in mitochondrial β-oxidation of long-chain FAs [46]. ESRRA is an important regulator of the acyl-CoA dehydrogenase medium chain (*ACADM*) gene [47], whose enzyme catalyses the initial rate limiting step in β-oxidation step of C4-C16 FAs with an optimum at C6-C8 [48]. The *PLA2G16* gene encodes a major regulator of lipolysis in adipose tissue and through the regulation of FA oxidation in adipocytes may change the FA profile [49]. Other candidate genes found in this region were insulin (*INS*) and insulin like growth factor 2 (*IGF2*). INS promotes FA uptake into cells and stimulates the expression of FA synthetic proteins [50]. In pigs, *IGF2* is responsible of 10-20% of the phenotypic variation in backfat thickness [51], and it has been associated with the C18:3(n-6) content in IMF [52]. The last two candidate genes of this region were patatin like phospholipase domain containing 2 (*PNPLA2*) and sirtuin 3 (*SIRT3*). The enzyme encoded by *PNPLA2* participates in the hydrolysis of stored triglycerides in adipose tissue [53], whereas the overexpression of one isoform of the *SIRT3* gene altered the FA composition in mouse skeletal muscle mitochondria including the MUFA/SFA ratio [54]. One or more candidate genes of SSC2 may be implicated in the genetic determination of the seven traits related with the FA composition in backfat. Therefore, further analysis of fine mapping would be needed to better elucidate the associations in this region.

The 80.45-82.57 Mb region of SSC4 was associated with the MUFA/PUFA ratio in backfat (Table 2). Located in an intergenic region, rs80848071 was the most significant SNP (p-value=6.01×10^{-6}) of this QTL. Two candidate genes were found inside this region: paired related homeobox 1 (*PRRX1*) and solute carrier family 19 member 2 (*SLC19A2*). PRRX1 is a transcription factor that negatively regulates adipogenesis in adipose tissue suppressing peroxisome proliferator

activated receptor gamma (*PPARG*) [55]. Therefore, *PPARG* suppression may increase MUFAs and decrease PUFAs in adipocytes [56]. The protein encoded by the other candidate gene, *SLC19A2*, is a thiamine transporter which has been associated with human type 2 diabetes mellitus [57]. In addition, free FAs in plasma were reduced in rats that were fed with a thiamine deficient diet [58].

The 15.34-27.05 Mb region of SSC6 was associated with the C20:3(n-6) abundance in backfat (Table 2). The rs81322046 SNP was the most significant (p-value=3.46×10^{-7}) and was located on an intergenic region. Four candidate genes were found in this region: glutamic-oxaloacetic transaminase 2 (*GOT2*), coenzyme 9 (*COQ9*), Bardet-Biedl syndrome 2 (*BBS2*), and solute carrier family 12 member 3 (*SLC12A3*). GOT2 is secreted from adipose tissue and is found in mitochondrion and cell surface facilitating uptake of long-chain free FAs [59]. In addition, GOT2 negatively regulates adipocyte differentiation [60]. COQ9 is also found in mitochondrion and acts as a lipid-binding protein playing an essential role for cellular respiration [61]. Finally, *BBS2* belongs to a family of genes that are involved in obesity [62], while mutations in *SLC12A3* affect human serum level of low-density lipoprotein cholesterol [63].

The 108.40-116.41 Mb region of SSC8 was associated with the FA content of C14:0, C16:0, and C16:1(n-7) in backfat (Table 2). Of the 14 significant SNPs found in this region, the most significant (p-value=3.05×10^{-10}) for the abundance of C16:0 in backfat was the rs81403349 variant, which was located on an intron of the ankyrin 2 (*ANK2*) gene. In this region the *ELOVL6* gene was mapped, a promising candidate gene involved in the elongation of even C12-C16 SFAs and MUFAs [64]. GWAS in different pig populations have found the association of this SSC8 QTL where the *ELOVL6* gene is located with the abundance of C16:0, C16:1(n-7), C18:1(n-9) in adipose tissue and muscle [21, 22, 24, 25]. In the BC1_LD animals, we have previously found a polymorphism in the promoter region of the *ELOVL6* gene (*ELOVL6:c.-533C>T)* strongly associated with the content of C16:0 and C16:1(n-7) in backfat and IMF [17]. Later on, the *ELOVL6:c.-394G>A* polymorphism was suggested as the causal mutation for the QTL on SSC8 that affects FA composition [33]. Other candidate genes were found in this region, *PLA2G12A* and hydroxyacyl-CoA dehydrogenase (*HADH*).

PLA2G12A liberates C20:4(n-6) from phospholipids [65], whereas HADH catalyses the oxidation of medium- and short-chain 3-hydroxy FAs [66].

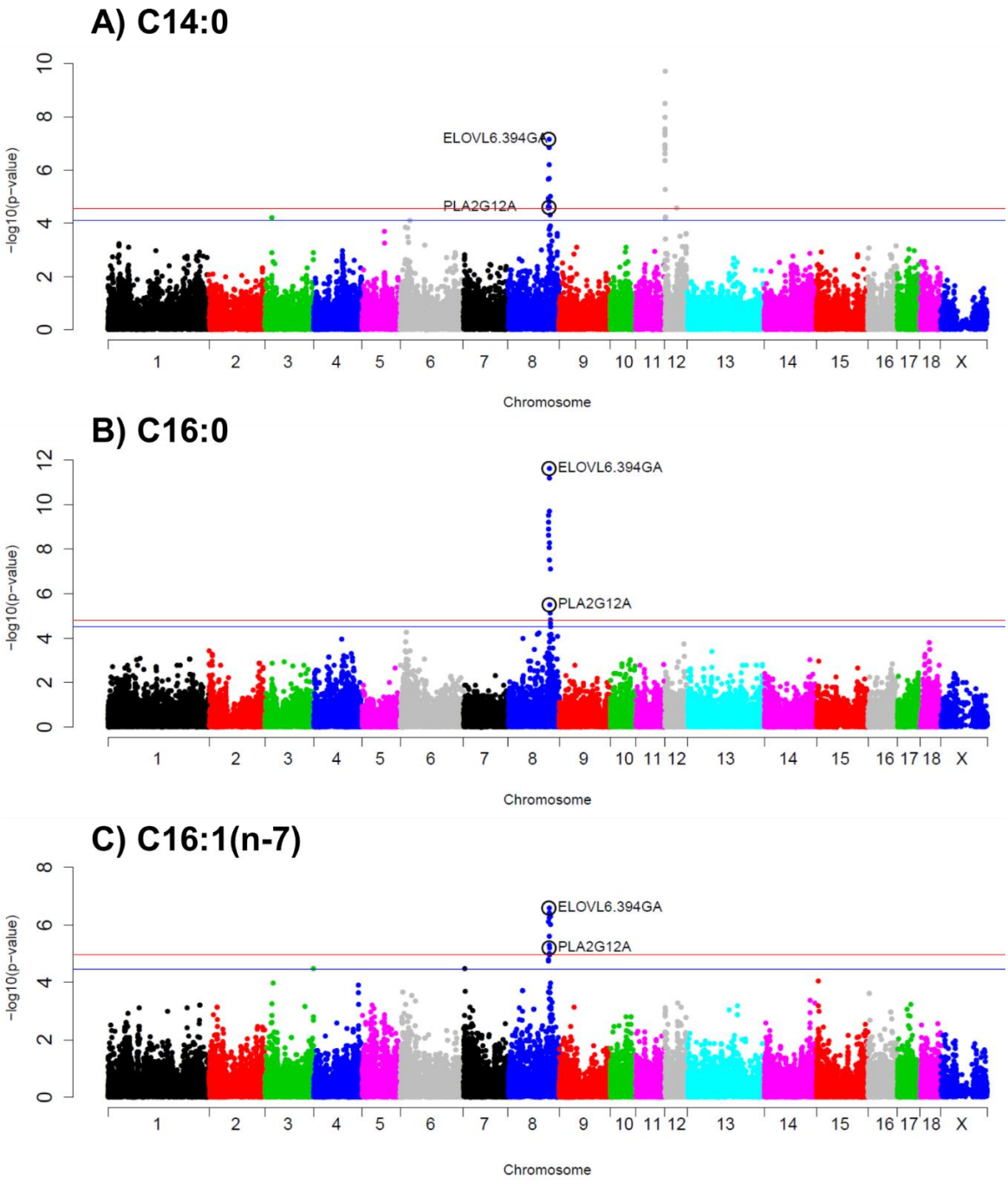


Figure 1. Manhattan plot representing the association analysis between the percentages of: C14:0 **(A)**, C16:0 **(B)** and C16:1(n-7) **(C)** in backfat and SNPs distributed along the pig genome, including eight new genotyped polymorphisms for SSC8. *ELOVL6:c.-394G>A* and *PLA2G12A* polymorphisms are included and labelled with a black circle. Red and blue lines indicate those SNPs that are below the genome-wide significance threshold (FDR ≤ 0.05 and FDR ≤ 0.1, respectively).

GWAS for the percentages of C14:0 (Figure 1A), C16:0 (Figure 1B), and C16:1(n-7) (Figure 1C) in backfat were reanalysed including seven SNPs of the *ELOVL6* gene and one SNP of the *PLA2G12A* gene. The seven SNPs located in *ELOVL6* have previously been described in the BC1_LD animals [17, 33], four of them were upstream variants of *ELOVL6* (rs325491325, rs345025813 (*ELOVL6:c.-533C>T*), rs341847499 and rs322956047 (*ELOVL6:c.-394G>A*)), one was located on the fourth exon (*ELOVL6:c.416C>T*), and two were located in the fifth exon (3'-UTR) (*ELOVL6:c.1408C>T* and *ELOVL6:c.1922C>T*). The *PLA2G12A* SNP (rs335191239) was an upstream variant that was found in the genome sequences of the BC1_LD founders, which are described in Revilla *et al*. [67]. In accordance with previous works in the BC1_LD pigs [33], *ELOVL6:c.-394G>A* was the most associated polymorphism with the percentages of C16:0 (*p*-value=2.43×10^{-12}) and C16:1(n-7) (*p*-value=2.6×10^{-7}) in backfat (Supplementary Table S3). The *ELOVL6:c.-394G>A* SNP was also the most associated polymorphism for the abundance of C14:0 (*p*-value=7.06×10^{-8}) in backfat, but this association has not been previously reported in the BC1_LD animals [33]. Apart from the *ELOVL6:c.-394G>A* polymorphism, other three *ELOVL6*-SNPs (*ELOVL6:c.-533C>T, ELOVL6:c.416C>T*, and *ELOVL6:c.1922C>T*) and the *PLA2G12A* SNP (rs335191239) were less significantly associated with the percentages of C14:0, C16:0, and C16:1(n-7) in backfat (Supplementary Table S1). In addition, *ELOVL6:c.-394G>A* was not segregating in the BC1_DU and BC1_PI backcrosses (Supplementary Table S4) and no region was found associated on SSC8 for the BC1_DU and BC1_PI pigs in the backcross-specific GWAS (Supplementary Table S5). The *p*-values obtained for the *ELOVL6:c.-394G>A* polymorphism in the backcross-specific GWAS performed in the BC1_LD were closer to the p-values obtained in the merged dataset GWAS for the percentages of C14:0 (2.41×10^{-8}), C16:0(9.93×10^{-15}), and C16:1(n-7) (2.00×10^{-8}) in backfat. Therefore, the *ELOVL6:c.-394G>A* polymorphism is probably the causal mutation of this QTL in BC1_LD for the C14:0 content in backfat as it has been previously proposed as the causal mutation for the percentages of C16:0 and C16:1(n-7) in backfat [33]. The A allele of the *ELOVL6:c.-394G>A* SNP in the BC1_LD animals was probably maternally inherited from the Landrace sows. In the BC1_LD, the G allele is the predominant

allele with a frequency of 0.63 and it is associated with a lower expression of the *ELOVL6* gene in comparison to the A allele [33]. Hence, a higher expression of *ELOVL6* gene produces the elongation of SFAs and MUFAs [64], increasing the content of long-chain FAs that may modify meat quality.

The 29.50-31.76 Mb region of SSC10 associated with the C16:1(n-9) abundance in backfat comprised three SNPs located in intergenic regions, being rs81423282 and rs81423288 (p-value=1.46×10^{-5}) the most significant SNPs (Table 2). The two candidate genes found in this region, neurotrophic receptor tyrosine kinase 2 (*NTRK2*) and RecQ mediated genome instability 1 (*RMI1*), are involved in obesity [68, 69]. However, further studies are required to discover how *NTRK2* or *RMI1* may be modifying the abundance in adipose tissue of a minor FA such as C16:1(n-9). Another region (51.87-53.92 Mb) on SSC10 was associated with the abundance of C20:3(n-6) in backfat (Table 2). The most significant SNP of this QTL was rs80979357 (p-value=7.46×10^{-6}), which is located on an intron of the *DNAJC1* (DnaJ heat shock protein family (Hsp 40) member C1) gene. *DNAJC1* is involved in FA synthesis [70] and has been found inside one genome-wide significant *locus* for subcutaneous adipose tissue in women [71]. On the other hand, another candidate gene located in this region was the *PIP4K2A* (phosphatidylinositol-5-phosphate 4-kinase type 2 alpha) gene. Liver of PIP4K2A-deficient mice were enriched in lipid droplets during fasting because autophagosomes failed to fuse with lysosomes at the rate needed [72]. Therefore, the impaired autophagy for recycling metabolites such as FAs may lead to cellular accumulation of C20:3(n-6) and other FAs in lipid droplets.

The 0-1.91 Mb region located at the beginning of SSC12 was associated with the C14:0 content in backfat (Table 2). Twelve SNPs were contained in this associated region and the most significant SNP was rs81308244 (p-value=1.98×10^{-10}), which was located in an intergenic region. In accordance with other GWAS performed in Erhualian and Duroc pigs [21, 23], a QTL signal for the C14:0 abundance in adipose tissue and muscle was detected on the same region of SSC12, where the *FASN* gene was located. FASN produces predominantly C16:0 and, to a lesser extent, C14:0 [73]. It is necessary to mention that along with *FASN*, the *ELOVL6* gene on SSC8 was also associated with the abundance of C14:0 in backfat, but *FASN* was not associated as *ELOVL6* with the

percentages of C16:0 and C16:1(n-7) (Figure 1). In addition, *FASN* and *ELOVL6* showed a higher expression in the adipose tissue of BC1_LD pigs with low PUFA content [27]. Although *FASN* is the most promising candidate gene associated with the variation of C14:0 content, it is necessary to mention that C14:0 is mainly taken from the diet [74] and it can be synthetized through other different pathways including C16:0 shortening [75]. In this sense, the proteins encoded by other candidate genes of this region may be affecting the C14:0 content in backfat. NOTUM (notum, palmitoleoyl-protein carboxylesterase) can bind to C14:1(n-5) and C16:1(n-7) [76]. In adipocytes, ASPSCR1 (ASPSCR1, UBX domain containing tether for SLC2A4) sequesters SLC2A4 (solute carrier family 2 member 4), also known as GLUT4, controlling glucose uptake [77], while METRNL (meteorin like, glial cell differentiation regulator) promotes lipid metabolism and insulin sensitization [78].

The 29.67-48.63 Mb region of SSC16 was associated with the differences in the C20:1(n-9)/C20:0 ratio in backfat (Table 2). A total of 34 SNPs associated with the C20:1(n-9)/C20:0 ratio in backfat were found in this region. The two most significant SNPs, rs81297480 and rs81458871 (p-value=3.20×10^{-8}), were intergenic variants. Four candidate genes were found in this region, *ELOVL7*, *GZMA* (granzyme A), *PIK3R1*, and *PLPP1* (phospholipid phosphatase 1). The *ELOVL7* gene is a strong candidate gene to explain the variation in C20:0 and C20:1(n-9) due to its protein function, which elongates C16-C20 FAs, with a preference toward C18 FAs [79]. In accordance with other studies, the *ELOVL7* gene has also been proposed as a candidate gene to explain the GWAS signals associated with the C20:0 abundance and several metabolic indices, such as C20:1(n-9)/C20:0, in the abdominal fat and IMF of different pig populations [19, 21, 22]. A higher expression of *GZMA* was observed in the mesenteric adipose tissue of beef cattle with low gain when compared with high gain animals [80]. PIK3R1 regulates glucose import [81] and in rat ovaries, the PIK3R1/AKT pathway has been involved in stearoyl-CoA desaturase 2 (*SCD2*) expression [82]. The protein encoded by *PLPP1* converts lipids such as phosphatidic acid and LPA to diacylglycerols [83].

GWAS for the C20:1(n-9)/C20:0 ratio in backfat was reanalysed including three SNPs of the *ELOVL7* gene and two SNPs of the *PIK3R1* gene (Figure 2). Of the three SNPs of the *ELOVL7* gene, one was an upstream variant (rs322657523) and the other two were located in the 5'-UTR (rs343494956 (*ELOVL7:c.-46A>G*)) and the 3'-UTR (tenth exon) (rs325490947 (*ELOVL7:c.*1432A>G*)). The SNP of the *PIK3R1* gene located on the second exon was a non-synonymous variant (rs331708297 (*PIK3R1:c.472A>G*)), while the other *PIK3R1*-SNP was an upstream variant (rs322671019). All the five SNPs (the three *ELOVL7*-SNPs and the two *PIK3R1*-SNPs) were found in the genome sequences of the BC1_LD founders, which are described in Revilla *et al*. [67]. The two *PIK3R1*-SNPs passed the significance threshold (Supplementary Table S3), but were not the most significantly associated SNPs (rs322671019 *p*-value=2.09×10^{-5} and rs331708297 *p*-value=5.76×10^{-5}). The SNP located in the tenth exon of *ELOVL7* (*ELOVL7:c.*1432A>G*) was the only significantly associated *ELOVL7*-SNP (*p*-value=1.8×10^{-7}), but other two SNPs showed a lower *p*-value, rs81297480 and rs81458871 (both *p*-values=3.2×10^{-8}). Therefore, the *ELOVL7:c.*1432A>G* SNP is unlikely to be the causal mutation of the SSC16 QTL for the variation of the C20:1(n-9)/C20:0 ratio in backfat. In addition, *ELOVL7:c.*1432A>G* was not segregating in the BC1_DU pigs and no AA individuals were observed in the BC1_PI pigs (Supplementary Table S4). The higher significance (*p*-value=1.8×10^{-7}) obtained in the merged dataset GWAS for the *ELOVL7:c.*1432A>G* SNP than the significance (*p*-value=7.05×10^{-6}) obtained in the BC1_LD-specific GWAS was probably due to the inclusion of the heterozygous BC1_PI individuals in the analysis of the merged dataset. Nonetheless, these findings suggest that the *ELOVL7* gene is a clear candidate to explain the differences in the C20:1(n-9)/C20:0 ratio and further studies will be required to find the causal mutation.

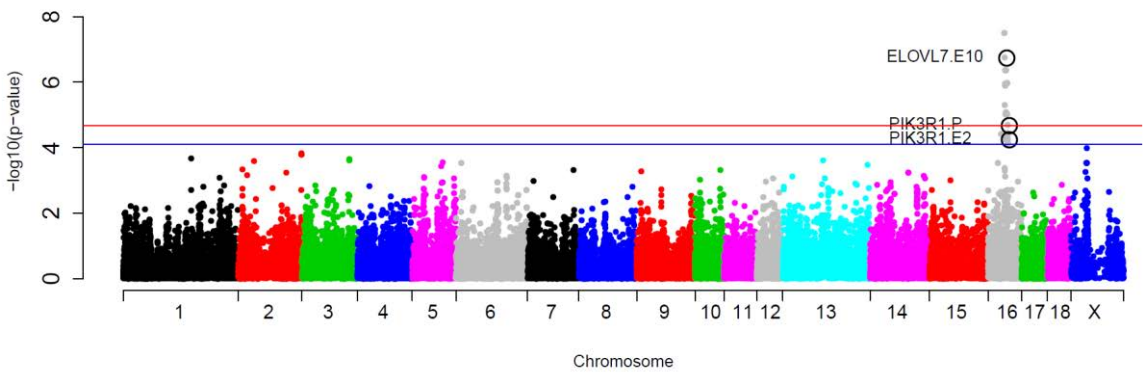


Figure 2. Manhattan plot representing the association analysis between the C20:1(n-9)/C20:0 ratio in backfat and SNPs distributed along the pig genome, including five new genotyped polymorphisms for SSC16. The *ELOVL7:c.*1432A>G*, *PIK3R1.P* (rs322671019) and *PIK3R1.E2* (rs331708297) polymorphisms are included and labelled with a black circle. Red and blue lines indicate those SNPs that are below the genome-wide significance threshold (FDR ≤ 0.05 and FDR ≤ 0.1, respectively).

QTLs for fatty acid composition in intramuscular fat and candidate genes

Two regions on SSC4 were associated with the FA composition of IMF. The first region (19.02-21.06 Mb) was associated with the C20:4(n-6)/C20:3(n-6) ratio (Table 3). The rs80910044 variant, located inside an intergenic region, was the most significant SNP (*p*-value=2.36×10^{-6}) of this QTL. Inside this region the ectonucleotide pyrophosphatase/phosphodiesterase 2 (*ENPP2*) gene was identified, also known as autotaxin (*ATX*). The protein encoded by *ENPP2* converts lysophosphatidylcholines to lysophosphatidic acids (LPAs) formed by different FAs such as C18:1(n-9) or C20:4(n-6) [84]. Furthermore, LPA is a signalling lipid involved in growth-factor like responses [85] and also participates in FA esterification being a precursor of the triglycerides that are deposited in IMF [86]. Other candidate genes in the SSC4 region that could be modulating the C20:4(n-6)/C20:3(n-6) ratio were exostosin glycosyltransferase 1 (*EXT1*) and nephroblastoma overexpressed (*NOV*). The mutation of hepatic *EXT1* increased the levels of plasma triglycerides in mice [87], whereas *NOV* was involved in the development of obesity [88] and the suppression of myogenesis [89].

The second region (122.76-124.98 Mb) on SSC4 was associated with the C20:3(n-3) content in IMF (Table 3). The two most significant SNPs (*p*-value=1.18×10^{-5}) of this QTL region were rs81347340 and rs80915252. While the

rs81347340 variant was located in an intergenic region, the rs80915252 variant was a splice region variant and a synonymous variant of the coiled-coil domain containing 18 (*CCDC18*) gene. Two candidate genes were found in this region: ATP binding cassette subfamily D member 3 (*ABCD3*) and glutamate-cysteine ligase modifier subunit (*GCLM*). *ABCD3* encodes for a protein that imports free FAs into peroxisomes where substrates are chain-shortened by β-oxidation [90]. In particular, ABCD3 has been reported to preferentially transport hydrophilic unsaturated FAs like C20:3(n-3) [90]. The other candidate gene, *GCLM*, plays a role in the metabolism of dietary lipids and mice with *GCLM*-deficiency were protected from weight gain and adipose deposition [91].

Three candidate genes were located in the 175.54-181.65 Mb region of SSC13 associated with the abundance of C20:3(n-3) in IMF (Table 3). The most significant variant found inside this region was the rs81441592 intergenic variant (p-value=1.81×10^{-6}). The lipase I (*LIPI*) gene encodes for a phospholipase that breaks down phosphatidic acid into LPA. In this sense, LIPI may have a similar role as the aforementioned ENPP2 in LPA production which would affect the FAs that are deposited in IMF [86]. The nuclear receptor interacting protein 1 (*NRIP1*) gene is involved in fat accumulation [92] and lipolysis [93]. The last candidate gene (*ssc-let-7c*) is transcribed into a microRNA, miR-let-7c, that regulates muscle growth in pigs, whereas other members of its family (miR-let-7a and miR-let-7e) regulate lipid deposition [94].

The first region (109.95-114.62 Mb) found on SSC14 was associated with two metabolic ratios in IMF, C18:1(n-9)/C18:0 and MUFA/SFA (Table 3). This QTL comprised a total of 21 significant associated SNPs, being the rs335655209 variant the most significant SNP (p-value=1.08×10^{-7}) for the C18:1(n-9)/C18:0 ratio as well as one of the most significant SNPs (p-value=1.14×10^{-6}) for the MUFA/SFA ratio. The rs335655209 variant was located in an intron of the *BTRC* (beta-transducin repeat containing E3 ubiquitin protein ligase) gene. These two desaturation ratios may be modulated by the *SCD* gene found in this region. SCD participates in the biosynthesis of C18:1(n-9) by desaturating C18:0 [95]. However, the *ELOVL3* (ELOVL fatty acid elongase 3) gene is also inside this region and may affect FA composition through the synthesis of C20-C24 SFAs and MUFAs [96]. In accordance with other studies, GWAS performed in pigs have

detected the association of the percentages of C16:0, C16:1(n-7), C18:0, C18:1(n-9), SFAs and MUFAs in adipose tissue and IMF with this SSC14 region where the *SCD* and *ELOVL3* genes are located [19–26]. Furthermore, polymorphisms in SCD have been related with differences in FA composition and desaturation ratios in swine IMF [97]. Therefore, *SCD* and *ELOVL3* are strong candidate genes to modulate the FA composition in muscle.

The second region (140.15-142.37 Mb) found on SSC14 was associated with the C18:0 abundance in IMF (Table 3). This region was comprised of six significant SNPs and three of them (rs318740977, rs80814938 and rs80883500) were the most significant variants (p-value=3.56×10^{-6}) for this QTL. The rs318740977 and rs80814938 SNPs were intronic variants of the *KNDC1* (kinase non-catalytic C-lobe domain containing 1) gene, whereas the rs80883500 SNP was located inside an intron of the CALY (calcyon neuron specific vesicular protein) gene. Three candidate genes were found inside this region: enoyl-CoA hydratase, short chain 1 (*ECHS1*); cytochrome P450 family 2 subfamily E member 1 (*CYP2E1*), and BCL2 interacting protein 3 (*BNIP3*). ECHS1 is involved in mitochondrial FA β-oxidation, but its activity is linked to short-chain FAs [98]. In the same manner, CYP2E1 has preference for short SFAs and long unsaturated FAs, showing no C18:0 hydroxylase activity [99]. In addition, CYP2E1 activity was inhibited by PUFAs but no by C16:0 and C18:0 [100]. Therefore, mutations in *ECHS1* or *CYP2E1* may increase the C18:0 abundance in IMF through the modification of short SFAs metabolism. On the other hand, *BNIP3* may be responsible of the differences in C18:0 as well. The *BNIP3* gene is a mitophagy regulator that, when silenced, suppressed *FASN*-mediated free FA synthesis [101].

The 30.06-32.87 Mb region of SSC17 was associated with the C20:0 content in IMF (Table 3). The most significant SNP of this region was rs324135473 (p-value=2.63×10^{-6}), located in an intergenic region. The acyl-CoA synthase short chain family member 1 (*ACSS1*) gene is located within this region. *ACSS1* was differentially expressed in bulls with extreme FA composition in muscle [102]. Therefore, in our material, ACSS1 may be increasing the SFA amount through the transformation of acetyl-CoA into FAs. Located in the SSC17 region was another relevant gene, pantothenate kinase (*PANK2*). Humans with mutations in *PANK2* present lower levels of some FAs compared to controls [103].

Furthermore, inside an intron of *PANK2*, it is located the *ssc-mir-103-2* gene, which is transcribed into the miR-103-2 microRNA. In adipocytes, miR-103 accelerates adipogenesis and increases the expression of lipid metabolism related genes such as fatty acid binding protein 4 (*FABP4*) and adiponectin, C1Q and collagen domain containing (*ADIPOQ*) [104]. The last candidate gene found in this SSC17 region that may be modulating the C20:0 content was abhydrolase domain containing 12 (*ABHD12*). ABHD12 has monoacylglycerol lipase activity and preferentially hydrolyses 2-arachidonoylglycerol, which is an ester of C20:4(n-6) and glycerol [105].

Comparison between merged dataset GWAS and backcross-specific GWAS

Backcross-specific GWAS for the FA composition in backfat found 19 associated regions in BC1_LD, seven in BC1_PI, and five in BC1_DU (Figure 3 and Supplementary Table S5), whereas the backcross-specific regions found for the FA composition in IMF were 36 in BC1_PI, 33 in BC1_LD, and 23 in BC1_DU (Figure 4 and Supplementary Table S5). Hence, more significant associated regions were found for the FA profile in IMF than in backfat. In backfat, all the regions for FA composition found in the merged dataset overlapped with at least one backcross-specific region. There was overlapping on the first region of SSC2 (0-10.77 Mb) between BC1_LD and BC1_PI for C18:2(n-6), PUFA and MUFA/PUFA. In addition, the 25.05-27.96 Mb region on SSC6 in backfat was shared between the BC1_DU and BC1_PI backcrosses for different FA composition traits (C14:0 and SFA, respectively). Similarly, all the intramuscular FA composition regions found in the merged dataset overlapped with at least one backcross-specific region, except for the last region of SSC14 (140.15-142.37 Mb). Two regions were associated with distinct traits in the three backcrosses: the 85.56-101.25 Mb region on SSC2 (MUFA/SFA on BC1_PI; C20:3(n-6), C20:4(n-6), PUFA, and PUFA/SFA on BC1_DU, and C20:3(n-3) on BC1_LD), and the 10.50-45.32 Mb region on SSC4 (C18:1(n-9), C18:2(n-6), MUFA, PUFA, MUFA/SFA, PUFA/SFA, and C20:2(n-6)/C18:2(n-6) on BC1_PI; C18:1(n-7)/C16:1(n-7) on BC1_DU, and C18:1(n-9), C18:2(n-6), MUFA, and MUFA/PUFA on BC1_LD). In addition, five regions were overlapping between two backcrosses on SSC9 (124.17-130.17 Mb between BC1_DU and BC1_PI), on SSC11 (46.32-58.27 Mb between BC1_DU and BC1_LD), a second region on SSC11 (61.64-

76.84 Mb between BC1_DU and BC1_PI), on SSC13 (174.35-191.27 Mb between BC1_DU and BC1_PI), and on SSC17 (44.22-58.14 Mb between BC1_LD and BC1_PI). Interestingly, there was overlapping between the regions on SSC8 for the FA composition (C16:0, C16:1(n-7), and C18:1(n-7)/C16:1(n-7)) in backfat (104.17-123.21 Mb) and in IMF (105.34-114.81 Mb) for the BC1_LD pigs, but the association with C14:0 was only found in backfat. In summary, using different genotypic backgrounds reduced the number of significant associated regions increasing the relevance and robustness of the detected ones. However, most of the associated regions found in the merged dataset were driven by one backcross and then, mixing backcrosses resulted in the loss of *loci* associated to a specific backcross.

Conclusions

Our results increase the knowledge of the genetic basis of FA composition and lipid metabolism. Our results describe regions that may be regulating lipid metabolism in backfat and intramuscular fat. In addition, polymorphisms of candidate genes such as *ELOVL6* and *ELOVL7* would be useful as genetic markers for meat quality selection of commercial pig breeds. However, some polymorphisms were breed-specific, and further analyses are required to find and evaluate possible causal mutations.

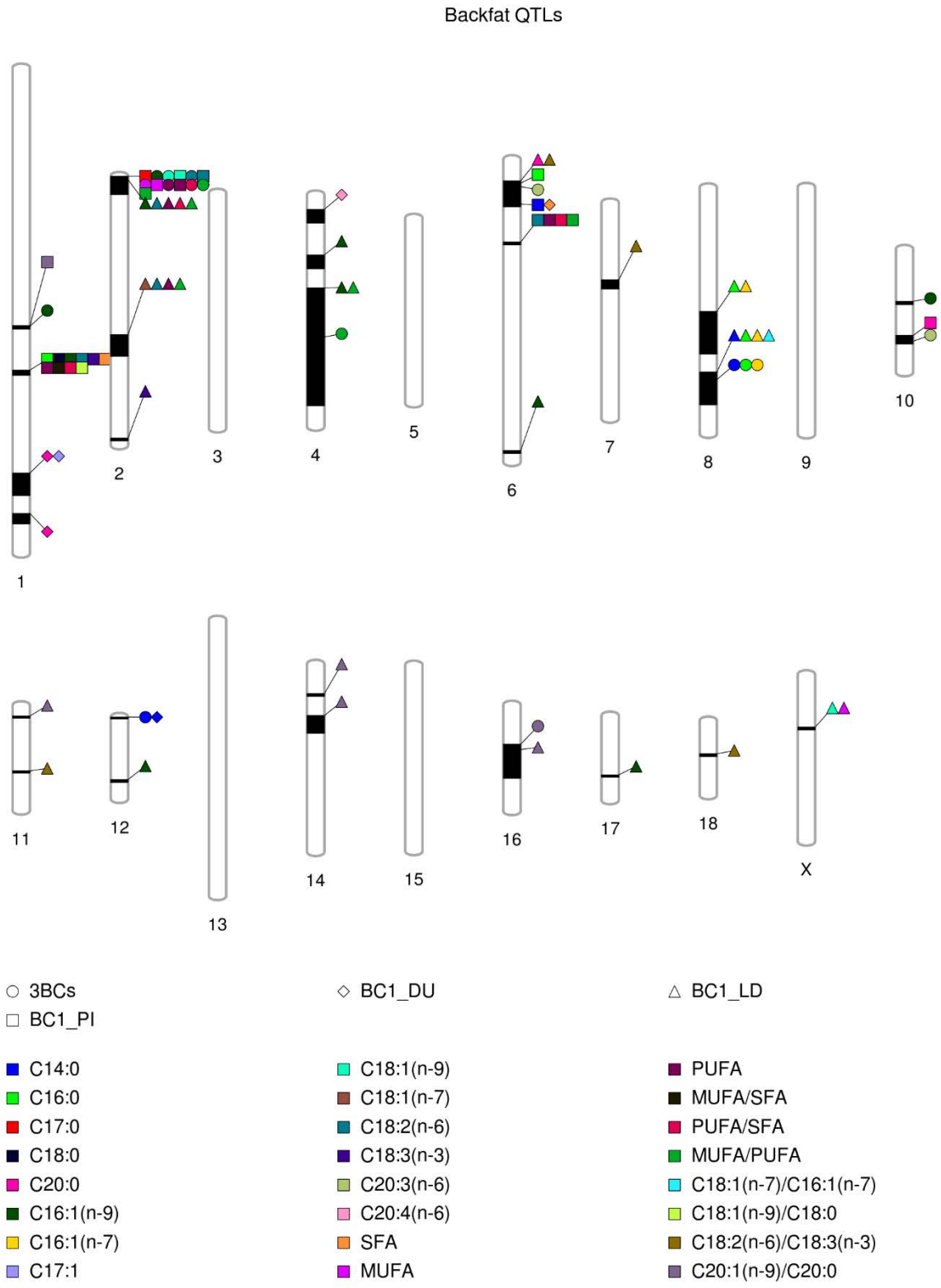


Figure 3. Comparison between the associated regions along pig chromosomes for backfat FA composition in the merged dataset and in each backcross individually. The shape indicates the backcross or the merged dataset and the colour indicates the phenotypic trait as it is indicated in the legend.

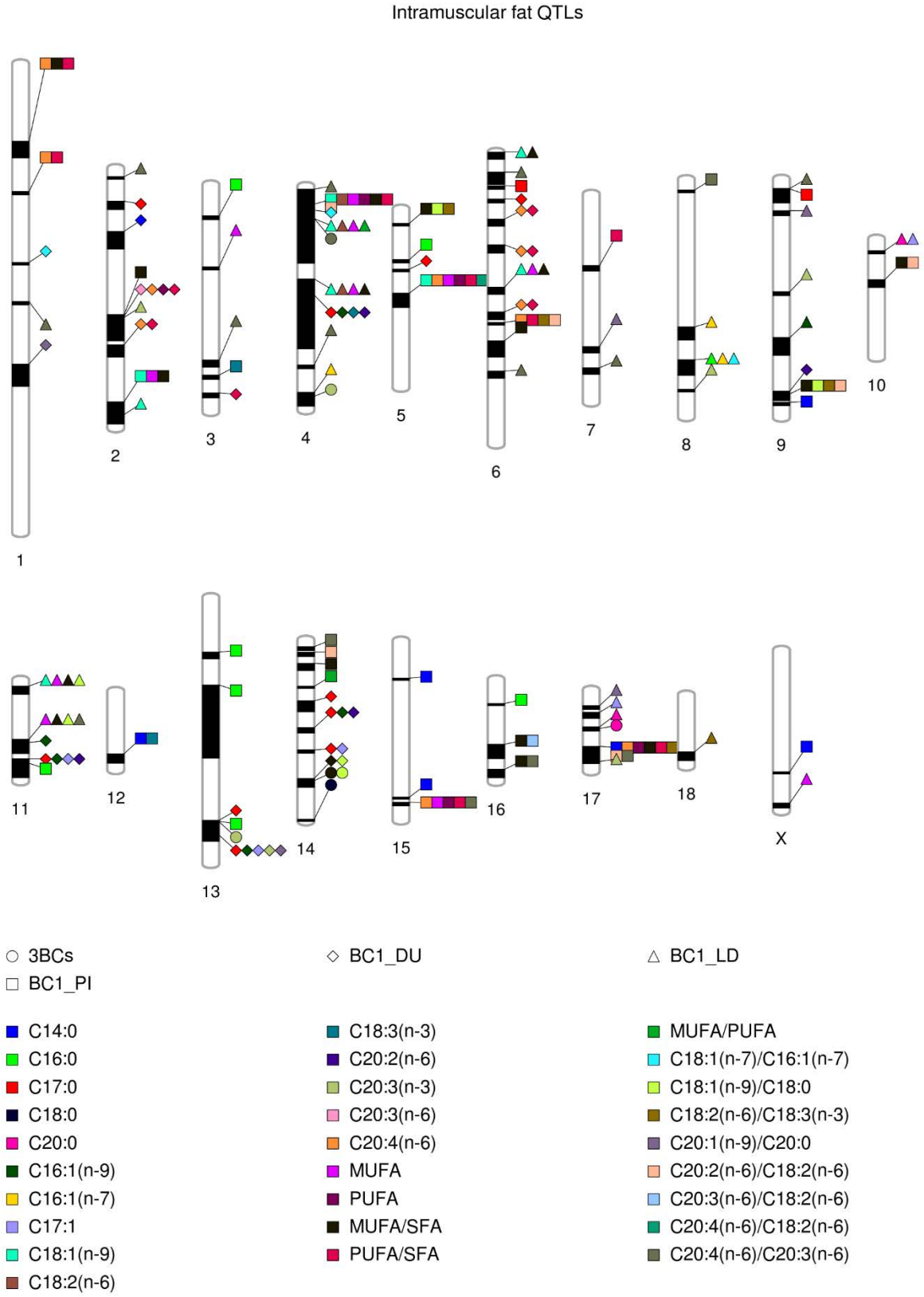


Figure 4. Comparison between the associated regions along pig chromosomes for intramuscular FA composition in the merged dataset and in each backcross individually. The shape indicates the backcross or the merged dataset and the colour indicates the phenotypic trait as it is indicated in the legend.

Declarations

Authors' contributions

JMF and AIF conceived and designed the experiments; JMF was the principal investigator of the project; this work is part of the PhD thesis of DCP co-supervised by MB and JMF; JLN led and supervised the generation of the material animal used in this work; JLN, AIF and JMF collected the animal samples; DCP, LCM and MR performed the pig genomic DNA extraction; AC genotyped the samples; DCP and JMF performed the genome-wide association studies; DCP, MB, and JMF wrote the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Spanish *Ministerio de Economía y Competitividad* (MINECO) and the *Fondo Europeo de Desarrollo Regional* (FEDER) with project references: AGL2014-56369-C2 and AGL2017-82641-R. D. Crespo-Piazuelo was funded by a "*Formació i Contractació de Personal Investigador Novell*" (FI-DGR) Ph.D grant from the *Generalitat de Catalunya* (ECO/1788/2014). L. Criado-Mesas was funded with a FPI grant from the AGL2014-56369-C2 project. M. Revilla was also funded by a FI-DGR (ECO/1639/2013). M. Ballester was financially supported by a "*Ramón y Cajal*" contract (RYC-2013-12573) from the Spanish *Ministerio de Economía y Competitividad*. We acknowledge the support of the Spanish *Ministerio de Economía y Competitividad* for the "*Severo Ochoa Programme for Centres of Excellence in R&D*" 2016-2019 (SEV-2015-0533) to the Centre for Research in Agricultural Genomics and the CERCA Programme / *Generalitat de Catalunya*.

References

1. Webb EC, O'Neill HA. The animal fat paradox and meat quality. Meat Sci. 2008;80:28–36.

2. Wood JD, Enser M, Fisher A V., Nute GR, Richardson RI, Sheard PR. Manipulating meat quality and composition. Proc Nutr Soc. 1999;58:363–70.

3. Wood JD, Enser M, Fisher A V., Nute GR, Sheard PR, Richardson RI, et al. Fat deposition, fatty acid composition and meat quality: A review. Meat Sci. 2008;78:343–58.

4. Mattson FH, Grundy SM. Comparison of effects of dietary saturated, monounsaturated, and polyunsaturated fatty acids on plasma lipids and lipoproteins in man. J Lipid Res. 1985;26:194–202.

5. Jiménez-Colmenero F, Ventanas J, Toldrá F. Nutritional composition of dry-cured ham and its role in a healthy diet. Meat Sci. 2010;84:585–93.

6. Bucher HC, Hengstler P, Schindler C, Meier G. N-3 polyunsaturated fatty acids in coronary heart disease: a meta-analysis of randomized controlled trials. Am J Med. 2002;112:298–304.

7. Cameron ND, Enser MB. Fatty acid composition of lipid in Longissimus dorsi muscle of Duroc and British Landrace pigs and its relationship with eating quality. Meat Sci. 1991;29:295–307.

8. Serra X, Gil F, Pérez-Enciso M, Oliver M., Vázquez J., Gispert M, et al. A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. Livest Prod Sci. 1998;56:215–23.

9. Ibáñez-Escriche N, Magallón E, Gonzalez E, Tejeda JF, Noguera JL. Genetic parameters and crossbreeding effects of fat deposition and fatty acid profiles in Iberian pig lines. J Anim Sci. 2016;94:28–37.

10. Kouba M, Sellier P. A review of the factors influencing the development of intermuscular adipose tissue in the growing pig. Meat Sci. 2011;88:213–20.

11. Benítez R, Fernández A, Isabel B, Núñez Y, De Mercado E, Gómez-Izquierdo E, et al. Modulatory Effects of Breed, Feeding Status, and Diet on Adipogenic, Lipogenic, and Lipolytic Gene Expression in Growing Iberian and Duroc Pigs. Int

J Mol Sci. 2017;19:22.

12. Cameron N. Genetic and phenotypic parameters for carcass traits, meat and eating quality traits in pigs. Livest Prod Sci. 1990;26:119–35.

13. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.

14. Ramos AM, Crooijmans RPM a, Affara N a, Amaral AJ, Archibald AL, Beever JE, et al. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One. 2009;4:e6524.

15. Groenen MAM. Development of a high-density Axiom® porcine genotyping array to meet research and commercial needs. In: Plant and Animal Genome XXIII Conference, San Diego, CA, USA. 2015.

16. Ramayo-Caldas Y, Mercadé A, Castelló A, Yang B, Rodríguez C, Alves E, et al. Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross. J Anim Sci. 2012;90:2883–93.

17. Corominas J, Ramayo-Caldas Y, Puig-Oliveras A, Pérez-Montarelo D, Noguera JL, Folch JM, et al. Polymorphism in the ELOVL6 Gene Is Associated with a Major QTL Effect on Fatty Acid Composition in Pigs. PLoS One. 2013;8:e53687.

18. Muñoz M, Rodríguez MC, Alves E, Folch JM, Ibañez-Escriche N, Silió L, et al. Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. BMC Genomics. 2013;14:845.

19. Yang B, Zhang W, Zhang Z, Fan Y, Xie X, Ai H, et al. Genome-wide association analyses for fatty acid composition in porcine muscle and abdominal fat tissues. PLoS One. 2013;8:e65554.

20. Ros-Freixedes R, Gol S, Pena RN, Tor M, Ibáñez-Escriche N, Dekkers JCM, et al. Genome-Wide Association Study Singles Out SCD and LEPR as the Two Main Loci Influencing Intramuscular Fat Content and Fatty Acid Composition in Duroc Pigs. PLoS One. 2016;11:e0152496.

21. Zhang W, Zhang J, Cui L, Ma J, Chen C, Ai H, et al. Genetic architecture of fatty acid composition in the longissimus dorsi muscle revealed by genome-wide association studies on diverse pig populations. Genet Sel Evol. 2016;48:5.

22. Zhang W, Bin Yang, Zhang J, Cui L, Ma J, Chen C, et al. Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations. Sci Rep. 2016;6:24718.

23. Sato S, Uemoto Y, Kikuchi T, Egawa S, Kohira K, Saito T, et al. Genome-wide association studies reveal additional related loci for fatty acid composition in a Duroc pig multigenerational population. Anim Sci J. 2017;88:1482–90.

24. van Son M, Enger EG, Grove H, Ros-Freixedes R, Kent MP, Lien S, et al. Genome-wide association study confirm major QTL for backfat fatty acid composition on SSC14 in Duroc pigs. BMC Genomics. 2017;18:369.

25. Zappaterra M, Ros-Freixedes R, Estany J, Davoli R. Association study highlights the influence of ELOVL fatty acid elongase 6 gene region on backfat fatty acid composition in Large White pig breed. Animal. 2018;48:1–10.

26. Viterbo VS, Lopez BIM, Kang H, Kim H, Song C, Seo KS. Genome wide association study of fatty acid composition in Duroc swine. Asian-Australasian J Anim Sci. 2018;31:1127–33.

27. Corominas J, Ramayo-Caldas Y, Puig-Oliveras A, Estellé J, Castelló A, Alves E, et al. Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. BMC Genomics. 2013;14:843.

28. Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, Hudson NJ, et al. Differences in Muscle Transcriptome among Pigs Phenotypically Extreme for Fatty Acid Composition. PLoS One. 2014;9:e99720.

29. Martínez-Montes ÁM, Fernández A, Muñoz M, Noguera JL, Folch JM, Fernández AI. Using genome wide association studies to identify common QTL regions in three different genetic backgrounds based on Iberian pig breed. PLoS One. 2018;13:e0190184.

30. Mach N, Devant M, Díaz I, Font-Furnols M, Oliver MA, García JA, et al. Increasing the amount of n-3 fatty acid in meat from young Holstein bulls through

nutrition. J Anim Sci. 2006;84:3039–48.

31. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press; 1989. p. E3–E4.

32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

33. Corominas J, Marchesi JAP, Puig-Oliveras A, Revilla M, Estellé J, Alves E, et al. Epigenetic regulation of the ELOVL6 gene is associated with a major QTL effect on fatty acid composition in pigs. Genet Sel Evol. 2015;47:20.

34. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44:821–4.

35. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57:289–300.

36. Wolfe D, Dudek S, Ritchie MD, Pendergrass SA. Visualizing genomic information across chromosomes with PhenoGram. BioData Min. 2013;6:18.

37. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford). 2011;2011 May:bar030.

38. Jurkowski W, Yazdi S, Elofsson A. Ligand binding properties of human galanin receptors. Mol Membr Biol. 2013;30:206–16.

39. Kim A, Park T. Diet-induced obesity regulates the galanin-mediated signaling cascade in the adipose tissue of mice. Mol Nutr Food Res. 2010;54:1361–70.

40. Kim S, Jin Y, Choi Y, Park T. Resveratrol exerts anti-obesity effects via mechanisms involving down-regulation of adipogenic and inflammatory processes in mice. Biochem Pharmacol. 2011;81:1343–51.

41. Li L, Wei S, Huang Q, Feng D, Zhang S, Liu Z. A novel galanin receptor 1a gene in zebrafish: tissue distribution, developmental expression roles in nutrition regulation. Comp Biochem Physiol B Biochem Mol Biol. 2013;164:159–67.

42. Guillou H, D'andrea S, Rioux V, Jan S, Legrand P. The surprising diversity of

Delta6-desaturase substrates. Biochem Soc Trans. 2004;32 Pt 1:86–7.

43. Rioux V, Choque B, Ezanno H, Duby C, Catheline D, Legrand P. Influence of the cis-9, cis-12 and cis-15 double bond position in octadecenoic acid (18:1) isomers on the rat FADS2-catalyzed Δ6-desaturation. Chem Phys Lipids. 2015;187:10–9.

44. Ballester M, Revilla M, Puig-Oliveras A, Marchesi JAP, Castelló A, Corominas J, et al. Analysis of the porcine APOA2 gene expression in liver, polymorphism identification and association with fatty acid composition traits. Anim Genet. 2016;47:552–9.

45. Mysore R, Zhou Y, Sädevirta S, Savolainen-Peltonen H, Nidhina Haridas PA, Soronen J, et al. MicroRNA-192* impairs adipocyte triglyceride storage. Biochim Biophys Acta. 2016;1861:342–51.

46. Prip-Buus C, Thuillier L, Abadi N, Prasad C, Dilling L, Klasing J, et al. Molecular and enzymatic characterization of a unique carnitine palmitoyltransferase 1A mutation in the Hutterite community. Mol Genet Metab. 2001;73:46–54.

47. Sladek R, Bader J a, Giguère V. The orphan nuclear receptor estrogen-related receptor alpha is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. Mol Cell Biol. 1997;17:5400–9.

48. Bross P, Engst S, Strauss AW, Kelly DP, Rasched I, Ghisla S. Characterization of wild-type and an active site mutant of human medium chain acyl-CoA dehydrogenase after expression in Escherichia coli. J Biol Chem. 1990;265:7116–9.

49. Ahmadian M, Duncan RE, Sul HS. The skinny on fat: lipolysis and fatty acid utilization in adipocytes. Trends Endocrinol Metab. 2009;20:424–8.

50. Saltiel AR, Kahn CR. Insulin signalling and the regulation of glucose and lipid metabolism. Nature. 2001;414:799–806.

51. Burgos C, Galve A, Moreno C, Altarriba J, Reina R, García C, et al. The effects of two alleles of IGF2 on fat content in pig carcasses and pork. Meat Sci. 2012;90:309–13.

52. Hong J, Kim D, Cho K, Sa S, Choi S, Kim Y, et al. Effects of genetic variants for the swine FABP3, HMGA1, MC4R, IGF2, and FABP4 genes on fatty acid composition. Meat Sci. 2015;110:46–51.

53. Lass A, Zimmermann R, Haemmerle G, Riederer M, Schoiswohl G, Schweiger M, et al. Adipose triglyceride lipase-mediated lipolysis of cellular fat stores is activated by CGI-58 and defective in Chanarin-Dorfman Syndrome. Cell Metab. 2006;3:309–19.

54. Chabi B, Fouret G, Lecomte J, Cortade F, Pessemesse L, Baati N, et al. Skeletal muscle overexpression of short isoform Sirt3 altered mitochondrial cardiolipin content and fatty acid composition. J Bioenerg Biomembr. 2018;50:131–42.

55. Du B, Cawthorn WP, Su A, Doucette CR, Yao Y, Hemati N, et al. The transcription factor paired-related homeobox 1 (Prrx1) inhibits adipogenesis by activating transforming growth factor-β (TGFβ) signaling. J Biol Chem. 2013;288:3036–47.

56. Sato D, Oda K, Kusunoki M, Nishina A, Takahashi K, Feng Z, et al. PPARγ activation alters fatty acid composition in adipose triglyceride, in addition to proliferation of small adipocytes, in insulin resistant high-fat fed rats. Eur J Pharmacol. 2016;773:71–7.

57. Tanaka T, Kono T, Terasaki F, Yasui K, Soyama A, Otsuka K, et al. Thiamine prevents obesity and obesity-associated metabolic disorders in OLETF rats. J Nutr Sci Vitaminol (Tokyo). 2010;56:335–46.

58. Liang CC. Metabolic changes in rats during developing thiamin deficiency. Biochem J. 1975;146:739–40.

59. Bradbury MW, Berk PD. Mitochondrial aspartate aminotransferase: direction of a single protein with two distinct functions to two subcellular sites does not require alternative splicing of the mRNA. Biochem J. 2000;345 Pt 3:423–7.

60. Challa TD, Straub LG, Balaz M, Kiehlmann E, Donze O, Rudofsky G, et al. Regulation of De Novo Adipocyte Differentiation Through Cross Talk Between Adipocytes and Preadipocytes. Diabetes. 2015;64:4075–87.

61. Lohman DC, Forouhar F, Beebe ET, Stefely MS, Minogue CE, Ulbrich A, et

al. Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme Q biosynthesis. Proc Natl Acad Sci U S A. 2014;111:E4697-705.

62. Forti E, Aksanov O, Birk RZ. Temporal expression pattern of Bardet-Biedl syndrome genes in adipogenesis. Int J Biochem Cell Biol. 2007;39:1055–62.

63. An C, Zhang K, Su X. SLC12A3 variants modulate LDL cholesterol levels in the Mongolian population. Lipids Health Dis. 2017;16:29.

64. Matsuzaka T, Shimano H. Elovl6: a new player in fatty acid metabolism and insulin sensitivity. J Mol Med (Berl). 2009;87:379–84.

65. Gelb MH, Valentin E, Ghomashchi F, Lazdunski M, Lambeau G. Cloning and recombinant expression of a structurally novel human secreted phospholipase A2. J Biol Chem. 2000;275:39823–6.

66. Jones PM, Bennett MJ. Clinical applications of 3-hydroxy fatty acid analysis by gas chromatography-mass spectrometry. Biochim Biophys Acta. 2011;1811:657–62.

67. Revilla M, Puig-Oliveras A, Castelló A, Crespo-Piazuelo D, Paludo E, Fernández AI, et al. A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. PLoS One. 2017;12:e0177014.

68. Suwa A, Yoshino M, Yamazaki C, Naitou M, Fujikawa R, Matsumoto S, et al. RMI1 deficiency in mice protects from diet and genetic-induced obesity. FEBS J. 2010;277:677–86.

69. Yeo GSH, Connie Hung C-C, Rochford J, Keogh J, Gray J, Sivaramakrishnan S, et al. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. Nat Neurosci. 2004;7:1187–9.

70. Nakatsuka A, Wada J, Iseda I, Teshigawara S, Higashio K, Murakami K, et al. Vaspin is an adipokine ameliorating ER stress in obesity as a ligand for cell-surface GRP78/MTJ-1 complex. Diabetes. 2012;61:2823–32.

71. Sung YJ, Pérusse L, Sarzynski MA, Fornage M, Sidney S, Sternfeld B, et al. Genome-wide association studies suggest sex-specific loci associated with

abdominal and visceral fat. Int J Obes (Lond). 2016;40:662–74.

72. Lundquist MR, Goncalves MD, Loughran RM, Possik E, Vijayaraghavan T, Yang A, et al. Phosphatidylinositol-5-Phosphate 4-Kinases Regulate Cellular Lipid Metabolism By Facilitating Autophagy. Mol Cell. 2018;70:531–544.e9.

73. Christie WW, Hunter ML, Clegg RA. The effects of cerulenin on lipid metabolism in vitro in cellular preparations from the rat. Biochim Biophys Acta. 1981;666:284–90.

74. Legrand P, Rioux V. The complex and important cellular and metabolic functions of saturated fatty acids. Lipids. 2010;45:941–6.

75. Rioux V, Catheline D, Legrand P. In rat hepatocytes, myristic acid occurs through lipogenesis, palmitic acid shortening and lauric acid elongation. Animal. 2007;1:820–6.

76. Kakugawa S, Langton PF, Zebisch M, Howell S, Chang T-H, Liu Y, et al. Notum deacylates Wnt proteins to suppress signalling activity. Nature. 2015;519:187–92.

77. Bogan JS, Hendon N, McKee AE, Tsao T-S, Lodish HF. Functional cloning of TUG as a regulator of GLUT4 glucose transporter trafficking. Nature. 2003;425:727–33.

78. Li Z-Y, Song J, Zheng S-L, Fan M-B, Guan Y-F, Qu Y, et al. Adipocyte Metrnl Antagonizes Insulin Resistance Through PPARγ Signaling. Diabetes. 2015;64:4011–22.

79. Naganuma T, Sato Y, Sassa T, Ohno Y, Kihara A. Biochemical characterization of the very long-chain fatty acid elongase ELOVL7. FEBS Lett. 2011;585:3337–41.

80. Lindholm-Perry AK, Cunningham HC, Kuehn LA, Vallet JL, Keele JW, Foote AP, et al. Relationships between the genes expressed in the mesenteric adipose tissue of beef cattle and feed intake and gain. Anim Genet. 2017.

81. Hara K, Yonezawa K, Sakaue H, Ando A, Kotani K, Kitamura T, et al. 1-Phosphatidylinositol 3-kinase activity is required for insulin-stimulated glucose transport but not for RAS activation in CHO cells. Proc Natl Acad Sci U S A.

1994;91:7415–9.

82. Moreau C, Froment P, Tosca L, Moreau V, Dupont J. Expression and regulation of the SCD2 desaturase in the rat ovary. Biol Reprod. 2006;74:75–87.

83. Roberts R, Sciorra VA, Morris AJ. Human type 2 phosphatidic acid phosphohydrolases. Substrate specificity of the type 2a, 2b, and 2c enzymes and cell surface activity of the 2a isoform. J Biol Chem. 1998;273:22059–67.

84. Tokumura A, Harada K, Fukuzawa K, Tsukatani H. Involvement of lysophospholipase D in the production of lysophosphatidic acid in rat plasma. Biochim Biophys Acta. 1986;875:31–8.

85. Jalink K, Hordijk PL, Moolenaar WH. Growth factor-like effects of lysophosphatidic acid, a novel lipid mediator. Biochim Biophys Acta. 1994;1198:185–96.

86. Jeong J, Kwon EG, Im SK, Seo KS, Baik M. Expression of fat deposition and fat removal genes is associated with intramuscular fat content in longissimus dorsi muscle of Korean cattle steers. J Anim Sci. 2012;90:2044–53.

87. Mooij HL, Bernelot Moens SJ, Gordts PLSM, Stanford KI, Foley EM, van den Boogert MAW, et al. Ext1 heterozygosity causes a modest effect on postprandial lipid clearance in humans. J Lipid Res. 2015;56:665–73.

88. Pakradouni J, Le Goff W, Calmel C, Antoine B, Villard E, Frisdal E, et al. Plasma NOV/CCN3 levels are closely associated with obesity in patients with metabolic disorders. PLoS One. 2013;8:e66788.

89. Sakamoto K, Yamaguchi S, Ando R, Miyawaki A, Kabasawa Y, Takagi M, et al. The nephroblastoma overexpressed gene (NOV/ccn3) protein associates with Notch1 extracellular domain and inhibits myoblast differentiation via Notch signaling pathway. J Biol Chem. 2002;277:29399–405.

90. van Roermund CWT, Ijlst L, Wagemans T, Wanders RJA, Waterham HR. A role for the human peroxisomal half-transporter ABCD3 in the oxidation of dicarboxylic acids. Biochim Biophys Acta. 2014;1841:563–8.

91. Kendig EL, Chen Y, Krishan M, Johansson E, Schneider SN, Genter MB, et al. Lipid metabolism and body composition in Gclm(-/-) mice. Toxicol Appl

Pharmacol. 2011;257:338–48.

92. Leonardsson G, Steel JH, Christian M, Pocock V, Milligan S, Bell J, et al. Nuclear receptor corepressor RIP140 regulates fat accumulation. Proc Natl Acad Sci U S A. 2004;101:8437–42.

93. Ho P-C, Chuang Y-S, Hung C-H, Wei L-N. Cytoplasmic receptor-interacting protein 140 (RIP140) interacts with perilipin to regulate lipolysis. Cell Signal. 2011;23:1396–403.

94. Wang Z, Li Q, Chamba Y, Zhang B, Shang P, Zhang H, et al. Identification of Genes Related to Growth and Lipid Deposition from Transcriptome Profiles of Pig Muscle Tissue. PLoS One. 2015;10:e0141138.

95. Paton CM, Ntambi JM. Biochemical and physiological function of stearoyl-CoA desaturase. Am J Physiol Endocrinol Metab. 2009;297:E28-37.

96. Zadravec D, Brolinson A, Fisher RM, Carneheim C, Csikasz RI, Bertrand-Michel J, et al. Ablation of the very-long-chain fatty acid elongase ELOVL3 in mice leads to constrained lipid storage and resistance to diet-induced obesity. FASEB J. 2010;24:4366–77.

97. Uemoto Y, Nakano H, Kikuchi T, Sato S, Ishida M, Shibata T, et al. Fine mapping of porcine SSC14 QTL and SCD gene effects on fatty acid composition and melting point of fat in a Duroc purebred population. Anim Genet. 2012;43:225–8.

98. Sharpe AJ, McKenzie M. Mitochondrial Fatty Acid Oxidation Disorders Associated with Short-Chain Enoyl-CoA Hydratase (ECHS1) Deficiency. Cells. 2018;7:46.

99. Adas F, Berthou F, Salaün J., Dréano Y, Amet Y. Interspecies variations in fatty acid hydroxylations involving cytochromes P450 2E1 and 4A. Toxicol Lett. 1999;110:43–55.

100. Yao H-T, Chang Y-W, Lan S-J, Chen C-T, Hsu JTA, Yeh T-K. The inhibitory effect of polyunsaturated fatty acids on human CYP enzymes. Life Sci. 2006;79:2432–40.

101. Lee HJ, Jung YH, Choi GE, Ko SH, Lee S-J, Lee SH, et al. BNIP3 induction

by hypoxia stimulates FASN-dependent free fatty acid production enhancing therapeutic potential of umbilical cord blood-derived human mesenchymal stem cells. Redox Biol. 2017;13 May:426–43.

102. Berton MP, Fonseca LFS, Gimenez DFJ, Utembergue BL, Cesar ASM, Coutinho LL, et al. Gene expression profile of intramuscular muscle in Nellore cattle with extreme values of fatty acid. BMC Genomics. 2016;17:972.

103. Leoni V, Strittmatter L, Zorzi G, Zibordi F, Dusi S, Garavaglia B, et al. Metabolic consequences of mitochondrial coenzyme A deficiency in patients with PANK2 mutations. Mol Genet Metab. 2012;105:463–71.

104. Xie H, Lim B, Lodish HF. MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. Diabetes. 2009;58:1050–7.

105. Navia-Paldanius D, Savinainen JR, Laitinen JT. Biochemical and pharmacological characterization of human α/β-hydrolase domain containing 6 (ABHD6) and 12 (ABHD12). J Lipid Res. 2012;53:2413–24.

Supplementary Tables

Supplementary Table S1. List of the significantly associated SNPs within QTL regions for the FA composition in backfat when GWAS were performed in the merged dataset and the predicted consequences of the SNPs.

Supplementary Table S2. List of the significantly associated SNPs within QTL regions for the FA composition in IMF when GWAS were performed in the merged dataset and the predicted consequences of the SNPs.

Supplementary Table S3. Summary of the GWAS results for the thirteen SNPs genotyped in the three backcrosses.

Supplementary Table S4. Frequencies of the thirteen SNPs genotyped in the three backcrosses.

Supplementary Table S5. Summary of the associated regions found in the three backcrosses when GWAS were performed individually.

Paper II

Indel detection from Whole Genome Sequencing data and association with lipid metabolism in pigs

Crespo-Piazuelo, D., Lourdes Criado-Mesas, L., Revilla, R., Castelló, A., Fernández, A. I., Folch, J. M., and Ballester, M.

(*in preparation*)

Indel detection from Whole Genome Sequencing data and association with lipid metabolism in pigs

Daniel Crespo-Piazuelo*†, Lourdes Criado-Mesas*, Manuel Revilla*†, Anna Castelló*†, Ana I Fernández‡, Josep M Folch*† and Maria Ballester§

* Plant and Animal Genomics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, 08193 Bellaterra, Spain

† Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Spain

‡ Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), 28040 Madrid, Spain

§ Departament de Genètica i Millora Animal, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), 08140 Caldes de Montbui, Spain

Address for correspondence

D. Crespo-Piazuelo, Plant and Animal Genomics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, Edifici CRAG, Campus UAB, 08193 Bellaterra, Spain

E-mail: daniel.crespo@cragenomica.es

Summary

The selection in commercial swine breeds for meat-production efficiency has been increasing among the past decades, reducing the intramuscular fat content, which has changed the sensorial and technological properties of pork. Pig breeds are genetically divergent, being base pair substitutions (SNPs) the most studied polymorphisms. However, other variants like insertions and deletions (indels) should also be considered for animal breeding. Indels can have higher impact on the phenotypic variance than SNPs through frameshift mutations. In the present study, indel detection was performed from Whole Genome Sequencing data of Iberian boars and Landrace sows, using three different programs: *Dindel*, *SAMtools mpileup,* and *GATK*. A total of 1,928,746 indels were found in common with the three programs. The *VEP* tool predicted that 1,289 indels may have a high impact on protein sequence and function. Ten indels inside genes related with lipid metabolism were genotyped in pigs from three different backcrosses with Iberian origin, obtaining different allelic frequencies on each backcross. Genome-Wide Association Studies performed in the *Longissimus dorsi* muscle found an association between an indel located in the C1q and TNF related 12 (*C1QTNF12*) gene and the amount of eicosadienoic acid (C20:2(n-6)).

Keywords: fatty acid composition, GWAS, indel, intramuscular, pig, WGS

Introduction

Pork is one of the world's most produced meat. Swine breeding has been developed in parallel to the increase and intensification of this productive sector. For the last decades, meat-production efficiency in commercial pig breeds has been notably improved by genetic selection with the unwanted drawback of reducing the pork sensorial and technological properties of meat. These modifications came from the intramuscular fat (IMF) content reduction and the alterations of the fatty acid (FA) composition (Karlsson *et al.*, 1993). Thus, selected breeds as Landrace have an efficient meat production with a rapid growth and leaner carcass, but the resulting meat has low IMF and high polyunsaturated FAs (PUFA) content (Estévez *et al.*, 2003). Nevertheless, some rustic pig breeds, like the Iberian, possess high IMF content with a higher proportion of monounsaturated FAs (MUFA), which is relevant for cured products like ham (Serra *et al.*, 1998). In addition, MUFA have a more oxidative stability than PUFA, producing a better meat taste (Wood *et al.*, 2008). In contrast, PUFA consumption, in particular omega-3, has the beneficial role of decreasing the total cholesterol concentration, while saturated FAs (SFA) increase the risk of cardiovascular diseases (Poudyal *et al.*, 2011; Michas *et al.*, 2014).

The genetic divergence between breeds is driven through the accumulation of mutations. Apart from base pair substitutions, genetic mutations can be produced by insertion, inversion, fusion, duplication or deletion of the DNA sequences. The development of Next Generation Sequencing (NGS) methods has allowed a fast way of detecting these genomic variants. So far, the most well-known variants studied with this method are the substitutions of single nucleotide polymorphisms (SNPs), representing almost the 80% over all the detected variants (Mullikin *et al.*, 2000; Dawson, 2001; Weber *et al.*, 2002). In contrast, other genetic polymorphisms like the insertions and deletions (indels) have been less studied. Studies in *Drosophila melanogaster* and *Caenorhabditis elegans* have shown that indels represent between 16% and 25% of all genetic polymorphisms in these species (Berger *et al.*, 2001; Wicks *et al.*, 2001). In addition, studies performed in humans and chimpanzees reported that indels instead of SNPs

were the major source of evolutionary change (Britten, 2002; Anzai *et al.*, 2003; Britten *et al.*, 2003).

As it has been described over the last decades, the most frequently found indel was the 1 base pair (bp) long (Ophir and Graur, 1997; Zhang and Gerstein, 2003). Moreover, a mechanism that favours the occurrence of deletions was proposed by de Jong & Rydén (1981), in which the loops formed by slipped mispairing after DNA strand breakage are trimmed off. In accordance with this mechanism, a major proportion of deletions than insertions was observed in the genome of 18 mammals, with the exception of the opossum, but pig was not included in this study (Fan *et al.*, 2007). In some recent analyses performed in pigs, using Whole Genome Sequencing (WGS) data, the 1 bp long indel has been detected as the most frequent indel, but the deletion/insertion ratios differ (L. Chen *et al.*, 2014; Molnár *et al.*, 2014; Kang *et al.*, 2015; Wang *et al.*, 2017).

Indels can produce frameshifts in the reading frame of genes or change the total number of amino acids of the proteins, but they can also affect gene expression levels. In pigs, indels were found to affect backfat thickness (Ren *et al.*, 2012) and fat deposition (Zang *et al.*, 2016) through the alteration of gene expression, underlining the importance of these variants for animal production.

The objectives of this study were to identify indels from WGS data of Iberian and Landrace pigs, which were founders of an experimental cross (IBMAP) with productive records for FA composition, and to study the association between a selection of indels and meat quality traits in three different genetic backgrounds.

Material and methods

Ethics Statement

The present study was performed in accordance with the regulations of the Spanish Policy for Animal Protection RD1201/05, which meets the European Union Directive 86/609 about the protection of animals used in experimentation. All experimental procedures followed national and institutional guidelines for the Good Experimental Practices and were approved by the IRTA (*Institut de Recerca i Tecnologia Agroalimentàries*) Ethics Committee.

Animal material and phenotypic records

The main breed pig lines used in this study were Iberian and Landrace. The Iberian line, called Guadyerbas, is a unique black hairless line that has been genetically isolated in Spain since 1945 (Pérez-Enciso *et al.*, 2000). The Landrace line belonged to the experimental farm Nova Genètica S.A. (Lleida, Spain). WGS data of seven founders of the IBMAP experimental population (Pérez-Enciso *et al.*, 2000), two Iberian boars and five Landrace sows, were used for indel detection. Analysis of indel segregation and association with meat quality traits were performed in 451 individuals of different backcrosses: 160 BC1_LD ((Iberian x Landrace) x Landrace), 148 BC1_DU ((Iberian x Duroc) x Duroc) and 143 BC1_PI ((Iberian x Pietrain) x Pietrain).

Animals were fed *ad libitum* with a cereal-based commercial diet and slaughtered at an average age of 179.8 ± 2.6 days with an average carcass weight of 72.2 kg. Blood samples from founder animals were collected and stored at -20ºC until analysis. Samples of diaphragm tissue were collected from backcrossed animals, snap-frozen in liquid nitrogen and stored at -80°C until analysis. Genomic DNA was extracted from all samples by the phenol-chloroform method (Sambrook *et al.*, 1989).

At the slaughterhouse, 200 g of *Longissimus dorsi* samples were collected from the three backcrosses. The IMF composition was measured with a protocol based on gas chromatography of methyl esters as described in Pérez-Enciso et al. (2000). In total, 20 traits were analysed: 17 intramuscular FAs and 3 FA metabolism indices (Table 1). Data values were normalized applying a $\log_2$ transformation when needed.

Table 1. Descriptive statistics including mean and SD of fatty acid composition and FA indices in the *Longissimus dorsi* muscle.

Group	Trait	Name	Mean	SD
SFA	C14:0	Myristic acid	1.17	0.24
	C16:0	Palmitic acid	22.99	1.51
	C17:0	Margaric acid	0.27	0.10
	C18:0	Stearic acid	14.21	1.44
	C20:0	Arachidic acid	0.26	0.11
MUFA	C16:1(n-9)	7-Hexadecenoic acid	0.36	0.11
	C16:1(n-7)	Palmitoleic acid	2.61	0.50
	C17:1	Heptadecenoic acid	0.23	0.10
	C18:1(n-9)	Oleic acid	37.08	5.78
	C18:1(n-7)	Vaccenic acid	3.91	0.33
PUFA	C20:1(n-9)	Gondoic acid	0.82	0.20
	C18:2(n-6)	Linoleic acid	11.92	5.01
	C18:3(n-3)	α-Linolenic acid	0.50	0.23
	C20:2(n-6)	Eicosadienoic acid	0.43	0.28
	C20:3(n-3)	Eicosatrienoic acid	2.53	2.04
	C20:3(n-6)	Dihomo-γ-linolenic acid	0.51	0.14
	C20:4(n-6)	Arachidonic acid	0.22	0.13
Metabolic Ratios	SFA	Saturated fatty acids	38.90	2.44
	MUFA	Monounsaturated fatty acids	44.78	6.25
	PUFA	Polyunsaturated fatty acids	15.88	7.26

Whole genome sequencing

The whole genome of seven founders of the IBMAP population was sequenced at CNAG (National Centre for Genome Analysis, Barcelona, Spain) on an Illumina HiSeq2000 instrument (Illumina®, USA). Paired-end sequencing libraries, with approximately 300 bp insert size, were generated using TruSeq DNA Sample Prep Kit (Illumina®, USA). For each sample, around 40 million 100 bp-long paired-end reads were produced with an average sequencing depth of 11.7x.

Sequences were trimmed based on their quality using the FastQC (Andrews, 2010) software. Then, reads were mapped against the reference genome

sequence assembly *Sscrofa10.2* using the Burrows-Wheeler Alignment (BWA) tool (Li and Durbin, 2009). Duplicated reads or those which were under a Phred-based quality score of 20 were removed. Finally, alignment result files (in bam format) were prepared for indel detection.

Indel detection and effects prediction

Several programs allow performing indel calling from WGS bam files. Following the article of Neuman et al. (2013) on the comparison of short indel detection programs, we applied the recommended pipelines on the use of these three programs: *Dindel* (version 1.01) (Albers *et al.*, 2011), *SAMtools mpileup* (version 0.1.19) (Li *et al.*, 2009), and *Genome Analysis Toolkit* (*GATK*) (version 3.4-46) (McKenna *et al.*, 2010).

The *Variant Effect Predictor* (*VEP*) (version 82) (McLaren *et al.*, 2010) tool of Ensembl (http://www.ensembl.org/) was used to quickly and accurately predict the effects and consequences of indels previously found on Ensembl-annotated transcripts (McLaren *et al.*, 2010). Furthermore, to predict the possible effect of an indel in the secondary structure of a protein, *JPred4* (Drozdetskiy *et al.*, 2015) was used.

Genotyping

For indel validation and association analysis, ten selected indels were genotyped in three experimental backcrosses: BC1_LD (n=160), BC1_DU (n=148), and BC1_PI (n=143) using Taqman OpenArray™ genotyping plates custom designed in a QuantStudio™ 12K flex Real-Time PCR System (ThermoFisher Scientific).

The same animals of BC1_LD and BC1_PI were genotyped with the Porcine SNP60K BeadChip (Illumina®), while BC1_DU samples genotypes were obtained with the Axiom Porcine Genotyping Array (Affymetrix®). A total of 38,424 SNPs was kept after selecting those variants shared by both genotyping platforms and removing SNPs with a minor allele frequency (MAF) less than 5% and SNPs with more than 5% missing genotype data.

Genome-Wide Association Analysis

Genome-Wide Association Studies (GWAS) were performed between the measured phenotypes of IMF composition and the previously genotyped variants of the three backcrosses (38,424 SNPs and nine indels) along the pig reference genome assembly *Sscrofa11.1*. In addition, specific GWAS were performed in each backcross individually. The studies were conducted with *GEMMA* (Zhou and Stephens, 2012) software following the mixed linear model:

$$y_{ijklm} = Sex_i + Batch_j + Backcross_k + \beta c_l + u_l + \lambda_l a_m + e_{ijklm} ,$$

where y_{ijklm} indicates the value of the phenotypic observation in the l^{th} individual; sex (two categories), batch (14 categories) and backcross (3 categories) are fixed effects; β is a covariate coefficient with c being carcass weight; u_l is the infinitesimal genetic random effect and distributed as $N(0, A\sigma_u)$, where A is the numerator of the kinship matrix; λ_l is a -1, 0, +1 indicator variable depending on the l^{th} individual genotype for the m^{th} SNP or indel; a_m represents the additive effect associated with the m^{th} SNP or indel; and e_{ijklm} is the random residual term. For the specific studies in each backcross, the fixed effect of the backcross was removed. The multiple test correction was conducted with the *p.adjust* function incorporated in R (www.r-project.org) using the false discovery rate (FDR) method developed by Benjamini and Hochberg (Benjamini and Hochberg, 1995). In order to consider a SNP or an indel as significant or suggestive a cut-off was set at FDR≤0.05 or FDR≤0.1, respectively.

Results and Discussion

Genome-wide detection of indels in Iberian and Landrace animals

Whole genome sequencing data of seven founders of the IBMAP population (two Iberian boars and five Landrace sows) were used for indel detection with *Dindel*, *SAMtools mpileup* and *GATK* software. *Dindel* was the program that detected the highest number of indels (3,380,221) as opposed to *SAMtools mpileup* and *GATK* (2,749,596 and 2,957,377, respectively). To reduce the rate of false positives, only indels (1,928,746) that were found in common between the three programs

were considered for further analyses (Figure 1). In addition, 50,528 indels that did not present the same genotype in at least two programs were discarded.

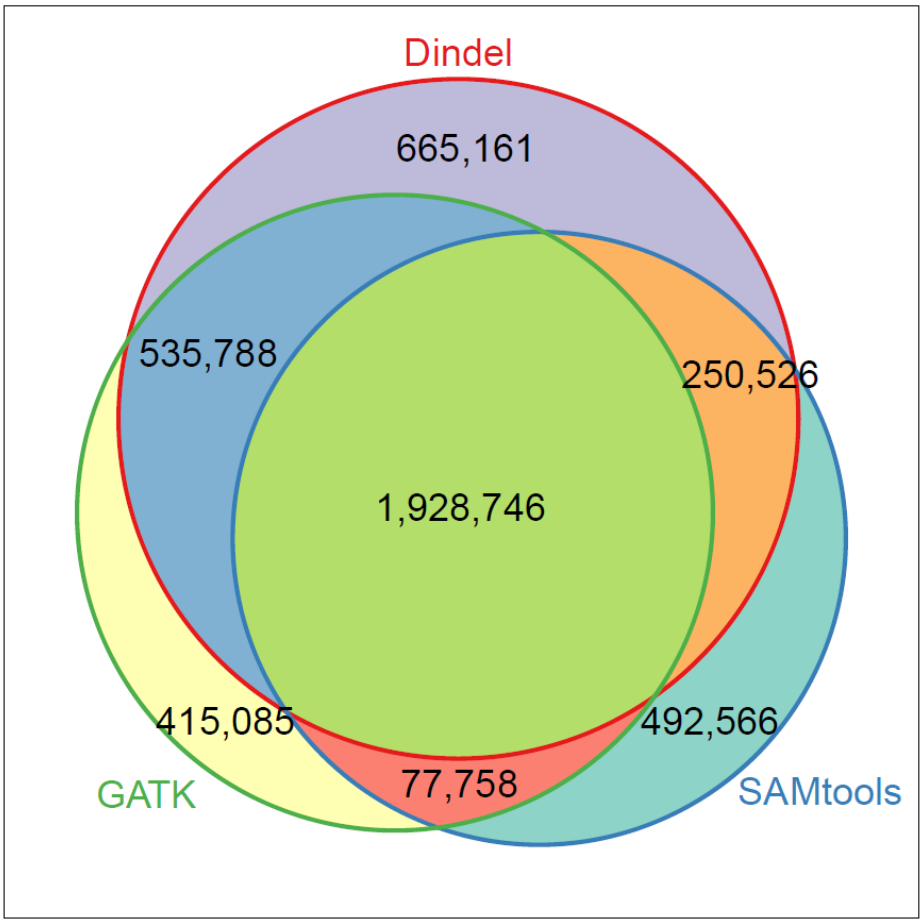


Figure 1. Weighted Venn diagram showing the number of indels shared between the three indel detection programs: *Dindel*, *Pindel* and *SAMtools mpileup*. A total of 1,928,746 indels were found in common.

Repetitive elements, such as microsatellites, are short insertions or deletions that can interfere with the detection and annotation of indels. Thus, to reduce the interference of repetitive elements in the next steps, 105,783 variants were discarded if they were triallelic or the alternative allele was different among individuals for the same chromosomal position. Moreover, 141,391 homozygous indels present in all samples were also trimmed considering that they were variants against the reference genome *Sscrofa10.2* and they may not segregate in our population. Hence, we only considered the final list comprising 1,631,044 indels for further analysis.

In a previous study of our group in which SNP calling was performed from WGS of these seven IBMAP founders, the number of SNPs identified after the quality filter was around 4.9 million in Iberian and 6 million in Landrace (unpublished data). Therefore, the number of indels detected (1.6 million indels) was within the expected range (16-25%) of the total number of variants detected (Mullikin *et al.*, 2000; Berger *et al.*, 2001; Dawson, 2001; Wicks *et al.*, 2001; Weber *et al.*, 2002). Nevertheless, another study in pigs reported that indels were less frequent than SNPs in a proportion of 1 to 10 (Molnár *et al.*, 2014).

The distribution of the quantity of the 1.6 million indels along all the *Sus scrofa* chromosomes was studied, showing that sexual chromosomes (SSCX and SSCY) had lower density of indels than autosomes (Figure 2). Disregarding the pseudoautosomic regions, this low density of indels in the sexual chromosomes is probably caused by the low recombination rate, only possible for the X chromosome in females, and by the appearance of hemizygous recessive lethal mutations in males. In autosomes, SSC10 had the highest density of indels, while SSC1 had the lowest (Figure 2).

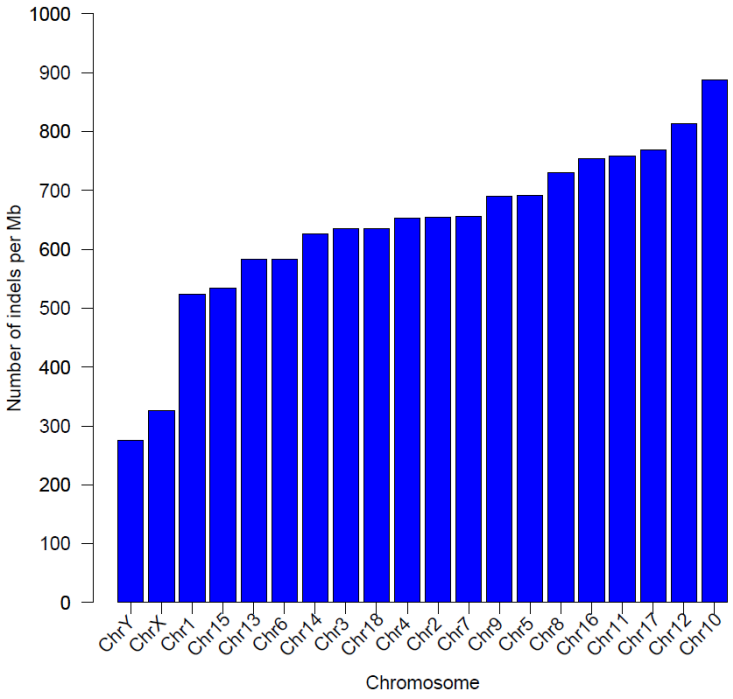


Figure 2. Distribution of the density of indels across chromosomes calculated as number of indels per Mb. Chromosomes are sorted in increasing order of density value.

In accordance with the literature, indel frequencies decreased as their length increased (Bentley *et al.*, 2008; L. Chen *et al.*, 2014) and thus, 1 bp long indel was the most frequent indel found (Figure 3), either insertion or deletion (Ophir and Graur, 1997; Zhang and Gerstein, 2003). Insertions were more frequent than deletions in single bp indels, but from the 1.6 million indels, 52.9% were deletions from 1 to 54 bp and the rest were insertions (47.1%) from 1 to 32 bp. Therefore, deletions were found to be more frequent than insertions, which has been previously reported by some other studies made in pigs (Molnár *et al.*, 2014; Kang *et al.*, 2015) and follows the mutational mechanisms described by de Jong & Rydén (1981).

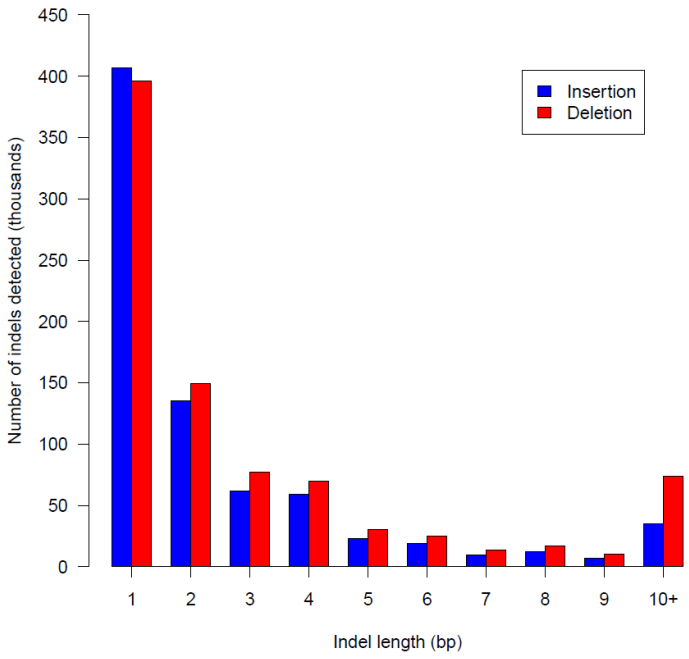


Figure 3. From the total of 1,631,044 indels detected, it is represented the quantity of them according to their length in bp. Insertions are in red and deletions are in blue.

Consequence and severity predictions of the indels detected

The effects (consequence type and severity) of the 1.6 million indels were estimated by the *VEP* platform and are summarized in Table 2. Since a variant may co-locate with more than one transcript, one line of output was provided for each instance of co-location and thus, there were more lines written (1,790,722) than indels entered (1,631,044). In addition, because of an indel could have more

than one effect in the same transcript (e.g., an indel could cause a frameshift along with a stop gained), the total number of separately predicted effects was 1,809,798. Around the third part of the 1.6 million indels (33.1%) did not fall within intergenic regions (539,920 indels) and only 1,758 indels were inside a coding region (0.11%). Finally, the *VEP* platform classified the 1.6 million indels by their possible severity as high (1,289), moderate (561) or low (1,018) impact, and the rest of indels were considered as modifiers.

Table 2. Consequences predicted by the *VEP* platform.

Consequence type	**Quantity**	***VEP* severity**
Total of indels processed (input)	**1,631,044**	-
intergenic variant	1,091,124	Modifier
intron variant	506,323	Modifier
downstream gene variant	91,517	Modifier
upstream gene variant	90,313	Modifier
non coding transcript variant	11,535	Modifier
3' UTR variant	7,618	Modifier
NMD transcript variant	5,911	Modifier
splice region variant	1,443	Low
frameshift variant	1,246	High
5' UTR variant	1,112	Modifier
non coding transcript exon variant	650	Modifier
inframe deletion	359	Moderate
inframe insertion	285	Moderate
coding sequence variant	115	Modifier
splice acceptor variant	94	High
splice donor variant	74	High
mature miRNA variant	41	Modifier
start lost	14	High
stop gained	9	High
protein altering variant	6	Moderate
stop retained variant	5	Low
stop lost	3	High
incomplete terminal codon variant	1	Low

Indel selection for genotyping

From the total of indels with high and moderate impact (1,850), ten indels were selected to be genotyped in three different genetic backgrounds. These indels were chosen regarding their possible consequence, if they were inside genes that could be related with lipid metabolism and/or considering their frequencies in the parental animals.

Table 3 summarizes the list of genes with indels selected for genotyping:

1. The aspartate beta-hydroxylase (*ASPH*) gene (ENSSSCG00000025087), located on SSC4, contained a predicted frameshift variant (rs691136075) with a high impact. The expression of this gene was found to be negatively correlated with insulin-stimulated sprouting in mice adipose tissue (Gealekman *et al.*, 2014).
2. The calpain 9 (*CAPN9*) gene (ENSSSCG00000010182) is located on SSC14 and contained a predicted inframe deletion (rs704351652). *CAPN9* is a member of the calpain family and some of its members have been associated with body fat content and insulin resistance in human and mice (Walder *et al.*, 2002; Cheverud *et al.*, 2010). This variant was found at extreme frequencies in the parental animals being the alternative allele (*CAPN9:c.2013_2015delGAA*) fixed in the Iberian boars.
3. The C-C motif chemokine receptor 7 (*CCR7*) gene (ENSSSCG00000017466) is located on SSC12 and contained a predicted frameshift variant (rs789030032). *CCR7* codifies for a chemokine receptor that plays a crucial role in inducing adipose tissue inflammation, insulin resistance and obesity (Sano *et al.*, 2015; Hellmann *et al.*, 2016). The allele frequency for this indel (*CCR7:c.1142dupA*) in the Landrace sows was 0.5 while the two Iberian boars were homozygous for the reference allele.
4. The C-reactive protein (*CRP*) gene (ENSSSCG00000021186), located on SSC4, contained a frameshift variant (*CRP:c.515delT*). High levels of *CRP* has been related with overweight and obesity in human adults (Visser, 1999). This variant was found fixed in the Iberian boars for the alternative allele (*CRP:c.515delT*) and the alleles of the Landrace sows were as the reference.

5. The C1q and TNF related 12 (*C1QTNF12*) gene (ENSSSCG00000003333) is located on SSC6 and contained an inframe deletion (*C1QTNF12:c.557_559delCCG*). This gene is also known as *CTRP12* and *FAM132A*. C1QTNF12 functions as an adipokine that is involved in glucose metabolism and obesity in mice (Enomoto *et al.*, 2011; Wei *et al.*, 2012a). This deletion was found at extreme frequencies in the founders being the alternative allele (*C1QTNF12:c.557_559delCCG*) fixed in the Iberian boars.
6. The granzyme A (*GZMA*) gene (ENSSSCG00000016903), located on SSC16, contained an inframe insertion (rs792025734). This gene was differentially expressed in the mesenteric adipose tissue of beef cattle with distinct gain (Lindholm-Perry *et al.*, 2017). The insertion (*GZMA:c.129_131dupGTT*) was found with a frequency of 0.8 in the Landrace sows while the Iberian boars were homozygous for the reference allele.
7. The jumonji domain containing 1C (*JMJD1C*) gene (ENSSSCG00000010226) is located on SSC14 and contained an inframe deletion (*JMJD1C:c.5964_5966delCAG*). *JMJD1C* was found in a human GWAS as a candidate gene for very low-density lipoprotein particles (Chasman *et al.*, 2009). This variation was found at extreme frequencies in the founders being the alternative allele (*JMJD1C:c.5964_5966delCAG*) fixed in the Iberian boars.
8. The lysosomal trafficking regulator (*LYST*) gene (ENSSSCG00000010151), located on SSC14, contained an inframe insertion (rs713515754). This gene has been related with hypertriglyceridemia and anomalous lipid and FA composition in the erythrocyte membranes of Chédiak-Higashi human patients (Chico *et al.*, 2000). This variation (*LYST:c.6287_6289dupCCA*) was found with a frequency of 0.8 in the Landrace sows while the Iberian boars were homozygous for the reference allele.
9. The peroxisomal biogenesis factor 19 (*PEX19*) gene (ENSSSCG00000023091) is located on SSC4 and contained a predicted frameshift variant (rs702520311). *PEX19* is assumed to be under

regulation by peroxisome proliferator-activated receptor gamma coactivator-1 alpha (PGC-1α) increasing the mitochondrial FA oxidation in human primary myotubes (Huang *et al.*, 2017). In addition, peroxisomes are intimately associated with lipid droplets and they are able to perform FA oxidation and lipid synthesis (Lodhi and Semenkovich, 2014). The frameshift variant was found to be fixed in the Iberian boars for the alternative allele (*PEX19:c.98_102dupAAGTC*), whereas in the Landrace sows the alternative allele was present with a frequency of 0.2.

10. The sterile alpha motif domain containing 4B (*SAMD4B*) gene (ENSSSCG00000016927), located on SSC16, contained a predicted frameshift variant that causes a stop gained (rs709630954). This gene was found to produce leanness and myopathy in mice due to the dysregulation of the rapamycin complex 1 (mTORC1) signalling (Z. Chen *et al.*, 2014).

Table 3. Selection of the ten genotyped indels with the alternative allele frequency in the Iberian (Freq. IB) and Landrace (Freq. LD) founders and their consequence predicted by the *VEP* platform.

Gene Name	Ensembl Gene ID	Chr	Position[a]	Reference Allele	Alternative Allele	Freq. IB	Freq. LD	Consequence	Severity
ASPH	ENSSSCG00000025087[b]	4	78,502,739	T	TAGAC	0	0.1	Frameshift variant Stop retained variant	High
PEX19	ENSSSCG00000023091	4	98,087,517	G	GCAAGT	1	0.2	Frameshift variant	High
CRP	ENSSSCG00000021186	4	98,755,304	GT	G	1	0	Frameshift variant	High
C1QTNF12	ENSSSCG00000003333	6	57,988,405	ACCG	A	1	0	Inframe deletion	Moderate
CCR7	ENSSSCG00000017466	12	22,151,183	T	TA	0	0.5	Frameshift variant Stop retained variant	High
LYST	ENSSSCG00000010151	14	59,597,172	G	GACC	0	0.8	Inframe insertion	Moderate
CAPN9	ENSSSCG00000010182	14	64,251,137	GTTC	G	1	0	Inframe deletion	Moderate
JMJD1C	ENSSSCG00000010226	14	71,899,504	TCTG	T	1	0	Inframe deletion	Moderate
GZMA	ENSSSCG00000016903	16	36,388,745	G	GTGT	0	0.8	Inframe insertion	Moderate
SAMD4B	ENSSSCG00000016927[b]	16	40,171,247	TCA	T	0.25	0.7	Frameshift variant Stop gained	High

[a]The position is referred to the indel's previous nucleotide in the *Sscrofa10.2* assembly.

[b]These indels were found in novel genes that were orthologous with the human, mouse and cow genes and the genotyped sequences were highly conserved with pig (data not shown).

Segregation analysis of the selected indels

The ten selected indels were genotyped in 160 BC1_LD, 148 BC1_DU and 143 BC1_PI individuals. Table 4 shows the genotype frequencies of indels in each backcross. Allele genotyping of the *CRP:c.515delT* indel failed and this indel was discarded for posterior analysis.

Table 4. Genotype frequencies of the nine indels found in each backcross. For each backcross, 148 BC1_DU, 160 BC1_LD and 143 BC1_PI were genotyped.

	ASPH			***PEX19***			***C1QTNF12***		
Alleles	*/*	ACAG/*	ACAG/ ACAG	*/*	AAGTC/*	AAGTC/ AAGTC	CCG/ CCG	CCG/*	*/*
BC1_DU	0.830	0.163	0.007	0.203	0.461	0.336	0.584	0.416	0
BC1_LD	0.645	0.316	0.039	0.025	0.247	0.728	0.529	0.471	0
BC1_PI	0.923	0.077	0	0	0.035	0.965	0.549	0.451	0
	CCR7			***LYST***			***CAPN9***		
Alleles	*/*	A/*	A/A	*/*	CCA/*	CCA/ CCA	GAA/ GAA	GAA/*	*/*
BC1_DU	0.616	0.336	0.048	0.717	0.221	0.062	0.423	0.577	0
BC1_LD	0.283	0.459	0.258	0.139	0.572	0.289	0.478	0.509	0.013
BC1_PI	0.851	0.142	0.007	0.289	0.556	0.155	0.538	0.455	0.007
	JMJD1C			***GZMA***			***SAMD4B***		
Alleles	CAG/ CAG	CAG/*	*/*	*/*	GTT/*	GTT/ GTT	TG/ TG	TG/*	*/*
BC1_DU	0.531	0.469	0	0.826	0.167	0.007	0.315	0.636	0.049
BC1_LD	0.513	0.487	0	0.170	0.446	0.384	0.051	0.503	0.446
BC1_PI	0.585	0.415	0	0.319	0.617	0.064	0.126	0.482	0.392

In accordance with the low frequency of the *ASPH:c.972_*1insACAG* indel in the Landrace founders (0.1), some *ASPH:c.972_*1insACAG* homozygotes were found in BC1_LD (0.039). In addition, a lower frequency of *ASPH:c.972_*1insACAG* homozygous individuals were found in BC1_DU (0.007), while no *ASPH:c.972_*1insACAG* homozygous individuals were found in BC1_PI. A similar allelic distribution was observed for the *CCR7:c.1142dupA* indel, with lower frequencies of *CCR7:c.1142dupA* homozygous individuals in BC1_DU and BC1_PI (0.048 and 0.007, respectively) than in BC1_LD (0.258). However, the high frequency of heterozygous individuals in BC1_LD (0.459) was probably a consequence of the high frequency of the *CCR7:c.1142dupA* allele in the Landrace founders (0.5).

Three indels located in the *C1QTNF12*, *CAPN9* and *JMJD1C* genes were found at extreme frequencies in the IBMAP founders. The alternative allele was fixed in the two Iberian boars, whereas the Landrace sows were homozygous for the reference allele. Then, no homozygous individuals for *C1QTNF12:c.557_559delCCG* and *JMJD1C:c.5964_5966delCAG* were found. This tendency was followed by the *CAPN9* indel, nonetheless, some *CAPN9:c.2013_2015delGAA* homozygous individuals were found in BC1_LD and BC1_PI (0.013 and 0.007, respectively). The frequencies observed in BC1_LD may be indicative that the *CAPN9:c.2013_2015delGAA*, *C1QTNF12:c.557_559delCCG*, and *JMJD1C:c.5964_5966delCAG* alleles were more common in the Iberian founders.

The *PEX19:c.98_102dupAAGTC* indel was also found at extreme frequencies in the IBMAP founders, being fixed in the Iberian boars and with a frequency of 0.2 in the Landrace sows. However, while homozygous individuals for the alternative alleles of the three previous indels (*C1QTNF12*, *CAPN9* and *JMJD1C*) were in a low frequency or absent in BC1LD, *PEX19:c.98_102dupAAGTC* homozygous individuals were present at a higher frequency (0.728). Meanwhile, the heterozygotes (0.462) were more abundant in the BC1_DU pigs and almost all the BC1_PI animals were *PEX19:c.98_102dupAAGTC* homozygous (0.965).

The indels of the *LYST* and *GZMA* genes, which were not present in the two founder Iberian boars and had a higher frequency of the alternative allele in the Landrace sows (0.8), were mostly found in heterozygosis in the BC1_LD and BC1_PI. In contrast, the BC1_DU had a higher proportion of homozygous individuals for the reference allele: 0.717 in the *LYST* gene and 0.826 in the *GZMA* gene.

Finally, the *SAMD4B*:*c.48_49delTG* indel was found in a distinct frequency from the previous ones in the IBMAP founders (0.25 for the Iberian and 0.7 for the Landrace). The genotyped frequencies in the three backcrosses were also very different, but most of the animals were heterozygotes (0.636 in the BC1_DU, 0.503 in the BC1_LD and 0.483 in the BC1_PI).

GWAS results

Nine indels located within genes related with lipid metabolism and genotyped in the three experimental backcrosses were selected for the association analysis. GWAS was performed with a linear-mixed model (GEMMA software) among the genotypes of 38,424 SNPs segregating in the three backcrosses and the nine selected indels and the fatty acid composition in muscle.

When GWAS were performed in the merged dataset, no association was found among the nine genotyped indels and the 20 FA composition traits in IMF. Then, backcross-specific GWAS were performed, finding a suggestive association between the *C1QTNF12:c.557_559delCCG* indel and the eicosadienoic acid (C20:2(n-6)) (*p*-value= 1.77×10^{-5}, FDR= 5.34×10^{-2}) in the BC1_PI animals (Figure 4).

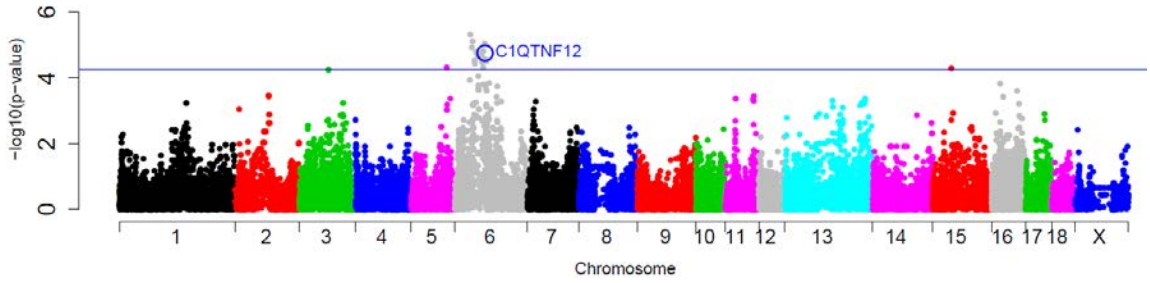


Figure 4. Manhattan plot representing the GWAS analysis for the relative abundance of eicosadienoic acid in the *Longissimus dorsi* muscle of the BC1_PI population where the *C1QTNF12* indel (blue circle) was suggestive (FDR≤0.1, blue line).

Eicosadienoic acid is the elongated product of linoleic acid, an essential FA that is taken from the diet (Lagarde *et al.*, 2013; Saini and Keum, 2018) and can be desaturated into arachidonic acid which participates in multiple regulatory pathways (Lagarde *et al.*, 2013; Saini and Keum, 2018). In the BC1_PI population, pigs with the *C1QTNF12:c.557_559delCCG* allele had a lower proportion of C20:2(n-6).

C1QTNF12 is a gene member of the C1QTNF family which preferentially acts in adipose tissue and liver regulating glucose uptake and fatty acid metabolism (Wei *et al.*, 2012a). C1QTNF12 can also form heterodimers with the protein encoded by the *ERFE* (erythroferrone) gene, another gene member of the C1QTNF family,

which is mainly expressed in skeletal muscle and is able to reduce the circulating levels of free FAs without affecting adipose tissue lipolysis (Seldin *et al.*, 2012). Therefore, alterations of the C1QTNF12/ERFE heterodimer may modify the circulation of free FAs and their accumulation in IMF.

Based on the Ensembl *Sscrofa11.1* genome assembly the porcine *C1QTNF12* gene consists of 8 exons and 7 introns (Ensembl ID: ENSSSCG00000003333). The identified indel produces an inframe deletion of three bases (CCG) in the exon 5 of *C1QTNF12*, which has the consequence of removing the alanine in the position 186 of the final protein. This alanine deletion was located in the C1q/TNF-like domain of C1QTNF12, a domain that is highly conserved among the *C1QTNF12* gene of mammals (Figure 5) and other vertebrate species (Wei *et al.*, 2012b), and is characteristic of the C1QTNF family. Furthermore, the alanine deletion in the position 186 was predicted to cause a new α-helix formation in the secondary structure of C1QTNF12, which could produce an impairment in the protein function (Figure 6).

However, the *C1QTNF12* indel was not the most significant genetic variant on SSC6 (Figure 4 and Table S1). Thus, further studies are required in order to analyse whether other genes or other *C1QTNF12* polymorphisms may be the cause for the differences in the eicosadienoic acid abundance.

In conclusion, in this study we used three different programs that allowed us to increase the accuracy of indel detection. Ten indels of the 1.9 million indels detected *in silico* were validated through genotyping in three different backcrosses, showing different allelic frequencies. In addition, a suggestive association was found between the *C1QTNF12:c.557_559delCCG* indel and the eicosadienoic acid abundance. Thus, indels can also be used as genetic markers associated with phenotypic traits of interest.

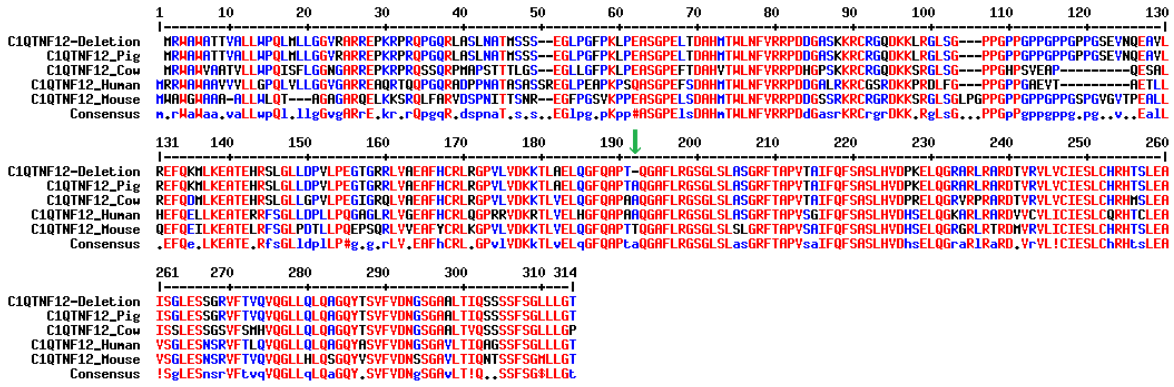


Figure 5. Multiple sequence alignment based on MULTALIN (Corpet, 1988) of the porcine C1QTNF12 protein sequence with the deletion and the reference sequences of the C1QTNF12 protein in pig, human, cow and mouse. The green arrow points out the deletion.

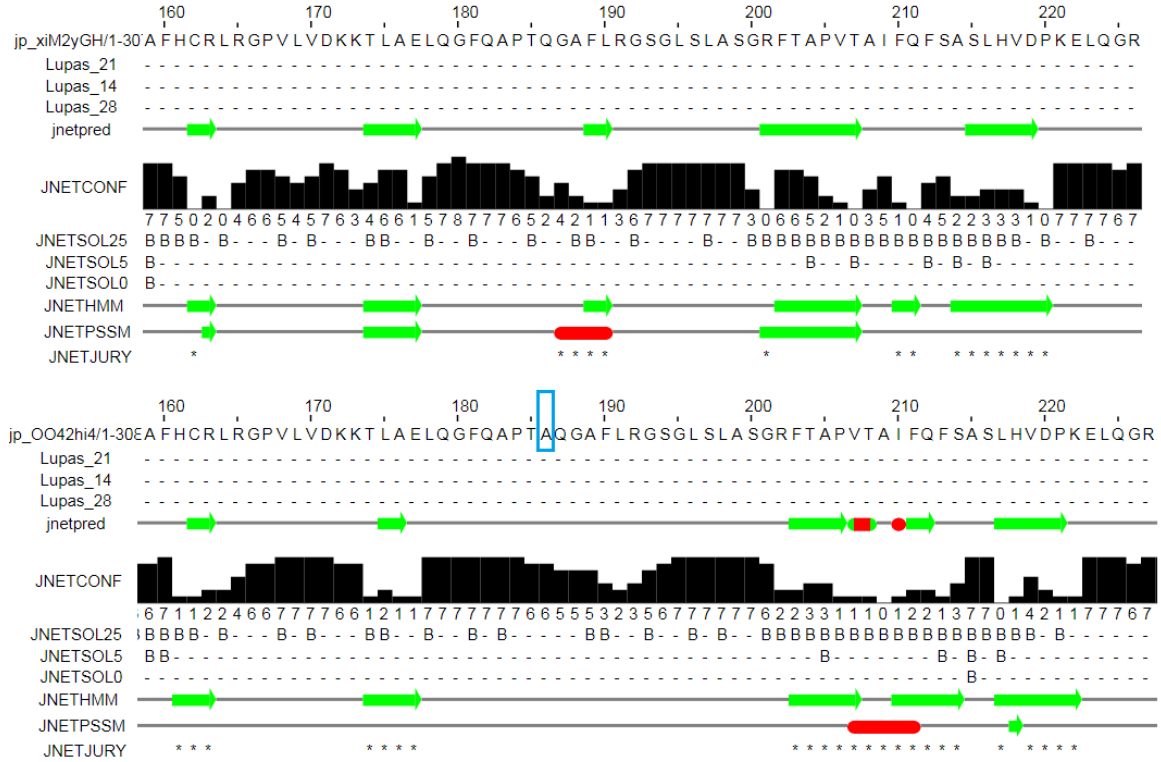


Figure 6. JPred4 prediction of the change in the secondary structure of the porcine C1QTNF12 protein when the alanine in the position 186 (A inside the blue rectangle) of the reference sequence (bottom) is deleted (above). Red segments represent alpha helices and green, beta sheets.

Acknowledgements

This work was funded by the *Ministerio de Economía y Competitividad* (MINECO) and the *Fondo Europeo de Desarrollo Regional* (FEDER) projects AGL2014-56369-C2-2-R and AGL2017-82641-R. D. Crespo-Piazuelo was funded by a "*Formació i Contractació de Personal Investigador Novell*" (FI-DGR) Ph.D grant from the *Generalitat de Catalunya* (ECO/1788/2014). L. Criado-Mesas was funded with a FPI grant from the AGL2014-56369-C2 project. M. Revilla was also funded by a FI-DGR (ECO/1639/2013). M. Ballester was financially supported by a "*Ramón y Cajal*" contract (RYC-2013-12573) from the Spanish Ministry of Economy and Competitiveness. We acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the "*Severo Ochoa Programme for Centres of Excellence in R&D*" 2016-2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics and the CERCA Programme / *Generalitat de Catalunya*.

References

Albers, C. a., Lunter, G., MacArthur, D. G., McVean, G., Ouwehand, W. H. and Durbin, R. (2011) 'Dindel: Accurate indel calls from short-read data', *Genome Research*, 21(6), pp. 961–973.

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data.*, *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Anzai, T., Shiina, T., Kimura, N., Yanagiya, K., Kohara, S., Shigenari, A., *et al.* (2003) 'Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(13), pp. 7708–13.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.', *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), pp. 289–300.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., *et al.* (2008) 'Accurate whole human genome sequencing using reversible terminator chemistry.', *Nature*, 456(7218), pp. 53–9.

Berger, J., Suzuki, T., Senti, K.-A., Stubbs, J., Schaffner, G. and Dickson, B. J. (2001) 'Genetic mapping with SNP markers in Drosophila.', *Nature genetics*, 29(4), pp. 475–81.

Britten, R. J. (2002) 'Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels', *Proceedings of the National Academy of Sciences*, 99(21), pp. 13633–13635.

Britten, R. J., Rowen, L., Williams, J. and Cameron, R. A. (2003) 'Majority of divergence between closely related DNA samples is due to indels', *Proceedings of the National Academy of Sciences*, 100(8), pp. 4661–4665.

Chasman, D. I., Paré, G., Mora, S., Hopewell, J. C., Peloso, G., Clarke, R., *et al.* (2009) 'Forty-Three Loci Associated with Plasma Lipoprotein Size, Concentration, and Cholesterol Content in Genome-Wide Analysis', *PLoS Genetics*. Edited by G. R. Abecasis, 5(11), p. e1000730.

Chen, L., Jin, L., Li, M., Tian, S., Che, T., Tang, Q., *et al.* (2014) 'Snapshot of Structural Variations in the Tibetan Wild Boar Genome at Single-Nucleotide Resolution', *Journal of Genetics and Genomics*, 41(12), pp. 653–657.

Chen, Z., Holland, W., Shelton, J. M., Ali, A., Zhan, X., Won, S., *et al.* (2014) 'Mutation of mouse Samd4 causes leanness, myopathy, uncoupled mitochondrial respiration, and dysregulated mTORC1 signaling.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), pp. 7367–72.

Cheverud, J. M., Fawcett, G. L., Jarvis, J. P., Norgard, E. A., Pavlicev, M., Pletscher, L. S., *et al.* (2010) 'Calpain-10 is a component of the obesity-related quantitative trait locus Adip1', *The Journal of Lipid Research*, 51(5), pp. 907–913.

Chico, Y., Lafita, M., Ramírez-Duque, P., Merino, F. and Ochoa, B. (2000) 'Alterations in erythrocyte membrane lipid and fatty acid composition in Chediak-Higashi syndrome.', *Biochimica et biophysica acta*, 1502(3), pp. 380–90.

Corpet, F. (1988) 'Multiple sequence alignment with hierarchical clustering.', *Nucleic acids research*, 16(22), pp. 10881–90.

Dawson, E. (2001) 'A SNP Resource for Human Chromosome 22: Extracting Dense Clusters of SNPs From the Genomic Sequence', *Genome Research*, 11(1), pp. 170–178.

Drozdetskiy, A., Cole, C., Procter, J. and Barton, G. J. (2015) 'JPred4: a protein secondary structure prediction server', *Nucleic Acids Research*, 43(W1), pp. W389–W394.

Enomoto, T., Ohashi, K., Shibata, R., Higuchi, A., Maruyama, S., Izumiya, Y., *et al.* (2011) 'Adipolin/C1qdc2/CTRP12 protein functions as an adipokine that improves glucose metabolism.', *The Journal of biological chemistry*, 286(40), pp. 34552–8.

Estévez, M., Morcuende, D. and Cava López, R. (2003) 'Physico-chemical characteristics of M. Longissimus dorsi from three lines of free-range reared Iberian pigs slaughtered at 90 kg live-weight and commercial pigs: a comparative study.', *Meat Science*, 64(4), pp. 499–506.

Fan, Y., Wang, W., Ma, G., Liang, L., Shi, Q. and Tao, S. (2007) 'Patterns of insertion and deletion in Mammalian genomes.', *Current genomics*, 8(6), pp. 370–8.

Gealekman, O., Gurav, K., Chouinard, M., Straubhaar, J., Thompson, M., Malkani, S., *et al.* (2014) 'Control of adipose tissue expandability in response to high fat diet by the insulin-like growth factor-binding protein-4.', *The Journal of biological chemistry*, 289(26), pp. 18327–38.

Hellmann, J., Sansbury, B. E., Holden, C. R., Tang, Y., Wong, B., Wysoczynski, M., *et al.* (2016) 'CCR7 Maintains Nonresolving Lymph Node and Adipose Inflammation in Obesity.', *Diabetes*, 65(8), pp. 2268–81.

Huang, T.-Y., Zheng, D., Houmard, J. A., Brault, J. J., Hickner, R. C. and Cortright, R. N. (2017) 'Overexpression of PGC-1α Increases Peroxisomal and Mitochondrial Fatty Acid Oxidation in Human Primary Myotubes', *American Journal of Physiology - Endocrinology And Metabolism*, p. ajpendo.00331.2016.

de Jong, W. and Rydén, L. (1981) 'Causes of more frequent deletions than insertions in mutations and protein evolution', *Nature*, 290(January), pp. 157–159.

Kang, H., Wang, H., Fan, Z., Zhao, P., Khan, A., Yin, Z., *et al.* (2015) 'Resequencing diverse Chinese indigenous breeds to enrich the map of genomic variations in swine.', *Genomics*. Elsevier Inc., 106(5), pp. 286–94.

Karlsson, A., Enfält, A. C., Essén-Gustavsson, B., Lundström, K., Rydhmer, L. and Stern, S. (1993) 'Muscle histochemical and biochemical properties in relation to meat quality during selection for increased lean tissue growth rate in pigs.', *Journal of animal science*, 71(4), pp. 930–8.

Lagarde, M., Bernoud-Hubac, N., Calzada, C., Véricel, E. and Guichardant, M. (2013) 'Lipidomics of essential fatty acids and oxygenated metabolites.', *Molecular nutrition & food research*, 57(8), pp. 1347–58.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760.

Lindholm-Perry, A. K., Cunningham, H. C., Kuehn, L. A., Vallet, J. L., Keele, J. W., Foote, A. P., *et al.* (2017) 'Relationships between the genes expressed in the mesenteric adipose tissue of beef cattle and feed intake and gain.', *Animal genetics*.

Lodhi, I. J. and Semenkovich, C. F. (2014) 'Peroxisomes: a nexus for lipid metabolism and cellular signaling.', *Cell metabolism*. Elsevier Inc., 19(3), pp. 380–92.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* (2010) 'The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, 20(9), pp. 1297–1303.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) 'Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor', *Bioinformatics*, 26(16), pp. 2069–2070.

Michas, G., Micha, R. and Zampelas, A. (2014) 'Dietary fats and cardiovascular disease: putting together the pieces of a complicated puzzle.', *Atherosclerosis*,

234(2), pp. 320–8.

Molnár, J., Nagy, T., Stéger, V., Tóth, G., Marincs, F. and Barta, E. (2014) 'Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary.', *BMC genomics*, 15(761), p. 761.

Mullikin, J. C., Hunt, S. E., Cole, C. G., Mortimore, B. J., Rice, C. M., Burton, J., *et al.* (2000) 'An SNP map of human chromosome 22.', *Nature*, 407(6803), pp. 516–520.

Neuman, J. a., Isakov, O. and Shomron, N. (2013) 'Analysis of insertion-deletion from deep-sequencing data: software evaluation for optimal detection', *Briefings in Bioinformatics*, 14(1), pp. 46–55.

Ophir, R. and Graur, D. (1997) 'Patterns and rates of indel evolution in processed pseudogenes from humans and murids.', *Gene*, 205(1–2), pp. 191–202.

Pérez-Enciso, M., Clop, A., Noguera, J. L., Ovilo, C., Coll, A., Folch, J. M., *et al.* (2000) 'A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross.', *Journal of animal science*, 78(10), pp. 2525–31.

Poudyal, H., Panchal, S. K., Diwan, V. and Brown, L. (2011) 'Omega-3 fatty acids and metabolic syndrome: effects and emerging mechanisms of action.', *Progress in lipid research*, 50(4), pp. 372–87.

Ren, Z., Liu, W., Zheng, R., Zuo, B., Xu, D., Lei, M., *et al.* (2012) 'A 304 bp insertion/deletion mutation in promoter region induces the increase of porcine IDH3β gene expression.', *Molecular biology reports*, 39(2), pp. 1419–26.

Saini, R. K. and Keum, Y. (2018) 'Omega-3 and omega-6 polyunsaturated fatty acids : Dietary sources , metabolism , and signi fi cance — A review', *Life Sciences*, 203(January), pp. 255–267.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) 'Molecular cloning: a laboratory manual.', in. Cold Spring Harbor Laboratory Press, pp. E3–E4.

Sano, T., Iwashita, M., Nagayasu, S., Yamashita, A., Shinjo, T., Hashikata, A., *et al.* (2015) 'Protection from diet-induced obesity and insulin resistance in mice lacking CCL19-CCR7 signaling.', *Obesity (Silver Spring, Md.)*, 23(7), pp. 1460–

71.

Seldin, M. M., Peterson, J. M., Byerly, M. S., Wei, Z. and Wong, G. W. (2012) 'Myonectin (CTRP15), a novel myokine that links skeletal muscle to systemic lipid homeostasis.', *The Journal of biological chemistry*, 287(15), pp. 11968–80.

Serra, X., Gil, F., Pérez-Enciso, M., Oliver, M. ., Vázquez, J. ., Gispert, M., *et al.* (1998) 'A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs', *Livestock Production Science*, 56(3), pp. 215–223.

Visser, M. (1999) 'Elevated C-Reactive Protein Levels in Overweight and Obese Adults', *JAMA*, 282(22), p. 2131.

Walder, K., McMillan, J., Lapsys, N., Kriketos, A., Trevaskis, J., Civitarese, A., *et al.* (2002) 'Calpain 3 gene expression in skeletal muscle is associated with body fat content and measures of insulin resistance.', *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity*, 26(4), pp. 442–9.

Wang, Z., Chen, Q., Liao, R., Zhang, Z., Zhang, X., Liu, X., *et al.* (2017) 'Genome-wide genetic variation discovery in Chinese Taihu pig breeds using next generation sequencing.', *Animal genetics*, 48(1), pp. 38–47.

Weber, J. L., David, D., Heil, J., Fan, Y., Zhao, C. and Marth, G. (2002) 'Human Diallelic Insertion/Deletion Polymorphisms', *The American Journal of Human Genetics*, 71(4), pp. 854–862.

Wei, Z., Peterson, J. M., Lei, X., Cebotaru, L., Wolfgang, M. J., Baldeviano, G. C., *et al.* (2012a) 'C1q/TNF-related Protein-12 (CTRP12), a Novel Adipokine That Improves Insulin Sensitivity and Glycemic Control in Mouse Models of Obesity and Diabetes', *Journal of Biological Chemistry*, 287(13), pp. 10301–10315.

Wei, Z., Lei, X., Seldin, M. M. and Wong, G. W. (2012b) 'Endopeptidase cleavage generates a functionally distinct isoform of C1q/tumor necrosis factor-related protein-12 (CTRP12) with an altered oligomeric state and signaling specificity.', *The Journal of biological chemistry*, 287(43), pp. 35804–14.

Wicks, S. R., Yeh, R. T., Gish, W. R., Waterston, R. H. and Plasterk, R. H. (2001) 'Rapid gene mapping in Caenorhabditis elegans using a high density

polymorphism map.', *Nature genetics*, 28(2), pp. 160–164.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., *et al.* (2008) 'Fat deposition, fatty acid composition and meat quality: A review.', *Meat science*, 78(4), pp. 343–58.

Zang, L., Wang, Y., Sun, B., Zhang, X., Yang, C., Kang, L., *et al.* (2016) 'Identification of a 13 bp indel polymorphism in the 3'-UTR of DGAT2 gene associated with backfat thickness and lean percentage in pigs.', *Gene*, 576(2 Pt 2), pp. 729–33.

Zhang, Z. and Gerstein, M. (2003) 'Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes', *Nucleic Acids Research*, 31(18), pp. 5338–5348.

Zhou, X. and Stephens, M. (2012) 'Genome-wide efficient mixed-model analysis for association studies.', *Nature genetics*, 44(7), pp. 821–4.

Additional information

Table S1. GEMMA output for the suggestive (FDR≤0.1) SNPs found in the GWAS analysis for the log2 normalization of the relative abundance of eicosadienoic acid in the *Longissimus dorsi* muscle of the BC1_PI population.

Paper III

Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions

Crespo-Piazuelo, D., Estellé, J., Revilla, M., Criado-Mesas, L., Ramayo-Caldas, Y., Óvilo, C., Fernández, A. I., Ballester, M., and Folch, J. M.

Scientific Reports (2018), 8(1), p. 12727

Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions

Daniel Crespo-Piazuelo[1,2,*], Jordi Estellé[3], Manuel Revilla[1,2], Lourdes Criado-Mesas[1], Yuliaxis Ramayo-Caldas[3,4], Cristina Óvilo[5], Ana I Fernández[5], Maria Ballester[4,] and Josep M Folch[1,2]

[1] Plant and Animal Genomics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, Bellaterra, Spain

[2] Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain

[3] Génétique Animale et Biologie Intégrative (GABI), Institut National de la Recherche Agronomique (INRA), AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

[4] Departament de Genètica i Millora Animal, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Caldes de Montbui, Spain

[5] Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain

*E-mail: daniel.crespo@cragenomica.es

Abstract

In addition to its value in meat production, the pig is an interesting animal model for human digestive tract studies due to its physiological similarities. The aim of this study was to describe the microbiome composition, distribution and interaction along the Iberian pig intestinal tract and its role in whole-body energy homeostasis. The V3-V4 region of the 16S rRNA gene was amplified and sequenced from the microbiomes of five gut sections (duodenum, jejunum, ileum, and proximal and distal colon) in thirteen castrated male pigs. A total of 1,669 operational taxonomic units distributed in 179 genera were found among all samples. The two most abundant genera in the small intestine were *Lactobacillus* and *Clostridium*, while *Prevotella* was predominant in the colon. The colon samples were more similar among the pigs and richer in species than the small intestine samples were. In the small intestine, the metagenome prediction pointed to rapid internalization and conversion of the available simple carbohydrates for microbial proliferation and maintenance. In the colon, a competition among anaerobic bacteria for plant polysaccharide degradation to produce short chain fatty acids was found. This study confirms that the energy pathways of the gut microbiome differ along its sections and provides a description of the correlations between genera.

Keywords: 16S rRNA gene, bacterial microbiota, intestinal tract, metagenome prediction, microbiome, OTUs, pig

Introduction

Trillions of microbes colonize the mammalian intestinal tract, supplying functions that in most cases the host cannot perform, such as digesting vegetable fibre and harvesting energy from otherwise inaccessible nutrients[1]. In humans, the genomes of these microorganisms (the so-called metagenome) contain more than 9 million unique genes[2]. With the publication of the first reference gene catalogue of the porcine gut microbiome[3], made through shotgun metagenome sequencing, the number of non-redundant genes identified reached 7.7 million in pigs. In addition to its interest for meat production, the pig is used as an animal model for human research due to the similarity in digestive tract anatomy, physiology, and immunology between pigs and humans[4]. Furthermore, both species share more non-redundant genes in their microbiota than humans do with other model organisms, such as the mouse[3]. Nevertheless, much of the work on the relationship between human obesity and the gut microbiota has been performed in mice[5].

Beyond whole-metagenome sequencing, an alternative cost-effective approach to studying the microbiota relies on targeted re-sequencing of the variable regions of the microbial 16S rRNA gene[6,7]. In recent years, the number of publications analysing the pig gut microbiota with the 16S approach has increased exponentially[8–14]. Interestingly, the pig gut microbiota composition has recently been related to average daily weight gain and to body weight[10,13], feed efficiency[15], feed conversion and feed intake[16]. However, only a few studies have analysed in detail the microbiota profiles in different parts of the digestive tract[17–19]. While these studies were focused on microbiota analysis along the digestive tract of the Large White[17], Laiwu[18] and Gloucestershire Old Spot[19] breeds, the microbiota profile of the Iberian pig and the correlations between genera along the pig gut have not yet been described.

The Iberian pig is a rustic animal with a higher adipogenic trend and lower meat efficiency than those of commercial breeds. Its high intramuscular fat content and backfat thickness are optimal for ham production. This excellent organoleptic quality is due to its high fat infiltration rate, with a high proportion of oleic fatty acid (C18:1(n-9)), along with a smaller proportion of polyunsaturated fatty acids[20].

In this context, Bäckhed *et al.*[21] found that, in mice, lipid metabolism can be modified by the gut microbiota. Therefore, unveiling the microbiota composition in Iberian pigs and how it varies along the digestive tract may provide a basis for understanding the effects of the microbiome on lipid metabolism in pigs.

The aim of this study was to describe the interactions and differences in the microbiome found along the Iberian pig gut and to evaluate their possible role in whole-body energy homeostasis. To this end, we explored the pig microbiota composition in five gut sections (duodenum, jejunum, ileum, and proximal and distal colon) by 16S rRNA gene sequencing.

Results

Microbial taxonomic composition shows major differences along the intestine

The V3-V4 region of the 16S rRNA was amplified and sequenced from the luminal contents of five gut sections (duodenum, jejunum, ileum, and proximal and distal colon) of thirteen Iberian pigs aged 120 days. A *MiSeq*® (*Illumina*®) instrument was used to obtain a mean of 126,549 sequences per sample. The sequences were processed and filtered through the QIIME pipeline[22], and a total of 1,669 operational taxonomic units (OTUs) were obtained among the five sections. According to the Greengenes 13.8 database[23], 643 were new OTUs. In addition, the 1,669 OTUs were aggregated into 179 genera and 18 phyla (Figure 1a). The two most abundant genera in the duodenum and jejunum were *Lactobacillus* (45.79% and 36.75%, respectively) and *Clostridium* (25.64% and 29.67%, respectively) (Figure 1b). Conversely, the two most abundant genera in the ileum were *Streptococcus* (17.73%) and the unspecified genera of the Clostridiaceae family (17.10%). Other genera of the Clostridiaceae family had a modest abundance in the ileum: *SMB53* (12.36%) and *Clostridium* (8.33%). The *Prevotella* genus was the most dominant in the colon, representing 40.90% in the proximal part and 34.99% in the distal one.

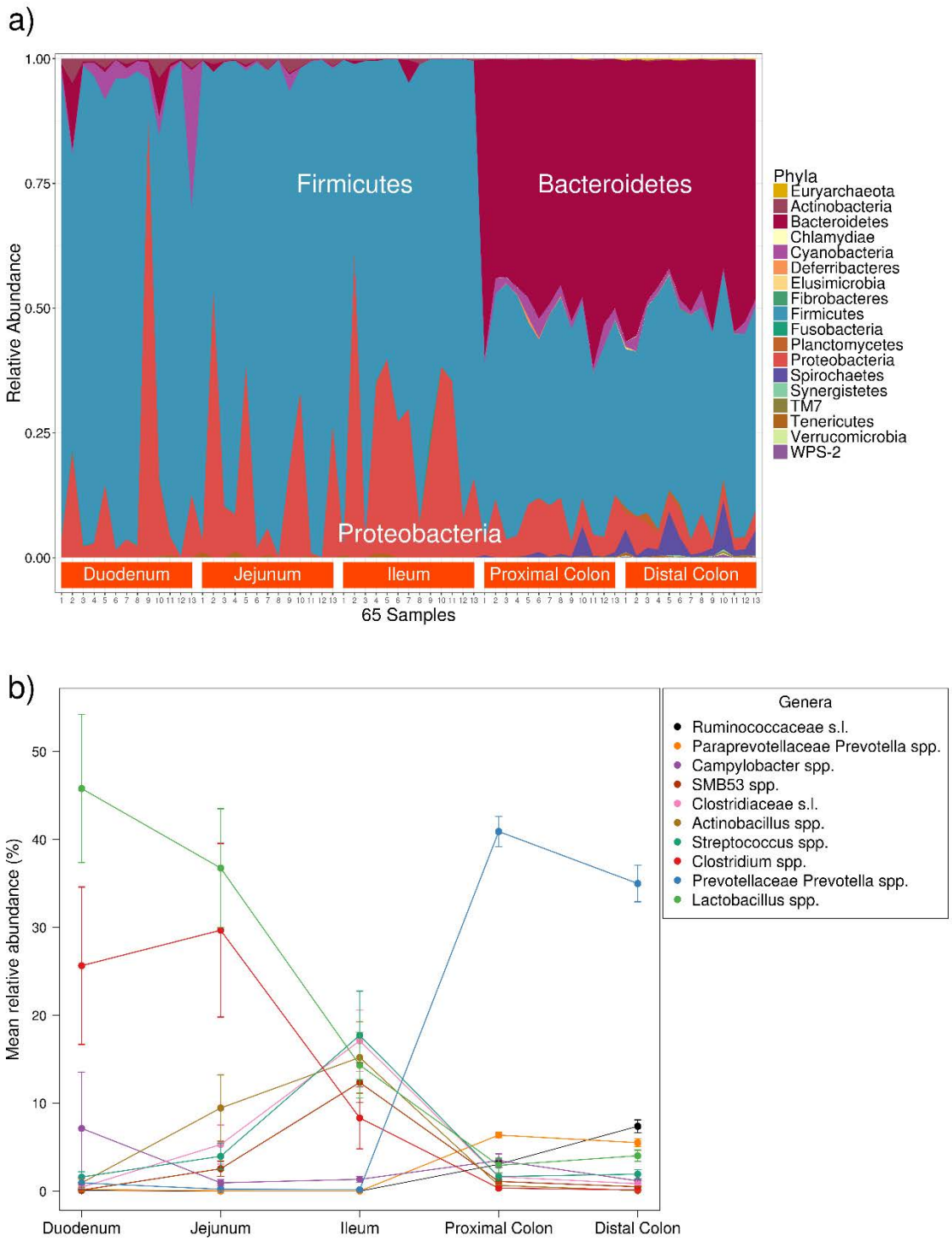


Figure 1. a) Stacked area plot of the OTUs grouped by phyla for the 65 samples sorted by intestinal section. **b)** Percentage evolution along the gut of the ten most abundant bacterial genera in the dataset. Segments represent the standard error.

Shannon index was used to evaluate the community α-diversity for each sample. The Shannon diversity measures the number of different species and their relative abundance within a sample. The α-diversity was higher in the large intestine than in the small intestine sections (Figure 2a). In fact, the small intestine samples showed a larger variation in α-diversity values among individuals, whereas the colon samples had higher and more constant values. On another level, the β-diversity measures the differences between samples. These differences are shown in Figure 2b, where the average pairwise distances of each group of samples to the group centroid (β-diversities) were obtained with the Whittaker index calculated through Bray-Curtis dissimilarities. In contrast with the α-diversity, which increased along the gut sections from the duodenum to the distal colon, the β-diversity decreased. Hence, there was a higher similarity among large intestine samples despite their high α-diversity, while small intestine regions such as the duodenum showed higher differences among individuals. These dissimilarities were observed in the non-metric multidimensional scaling (NMDS) plot, where samples were consistently separated according to their intestinal section (Figure 2c). Indeed, a large distance between the small and large intestines is depicted. The points representing the colon samples are closer to their centroid than those of the small intestine, in accordance with the β-diversity analysis (Figure 2b).

To describe which OTUs were present in all animals along all sections, that is, the minimum core microbiota, an OTU was assumed to be present if it had at least one count in a given sample (Figure 3). It is noticeable that the greatest number of OTUs were detected in only the distal colon samples. Moreover, the intersection between the two colonic regions had the highest number of OTUs. In contrast, the jejunum was the gut segment with the lowest number of unique OTUs. The duodenum, in accordance with its high β-diversity, also had a high number of unique OTUs in each sample. The intersection between the duodenum and the two large intestine regions shared a mean of 29.8 OTUs. However, the intersection between the duodenum and the other two small intestine sections shared a slightly greater number of OTUs (mean of 37.4). In addition, the number of OTUs shared among all the intestinal regions was between 27 and 40 for each animal. Nevertheless, when combining the datasets from all animals, 44 of the

1,669 OTUs were shared among the five intestinal regions, representing 71% of the total number of counts (see Supplementary Fig. S1 online). In the core formed by these 44 OTUs, only two were new OTUs, and the most abundant genera were *Lactobacillus* with 13 OTUs (23.74% of the core) and *Clostridium* with one OTU (22.45% of the core). Additionally, from the 1,669 total OTUs, 946 were absent in the small intestine sections, whilst 325 were not present in the large intestine (see Supplementary Fig. S1 online).

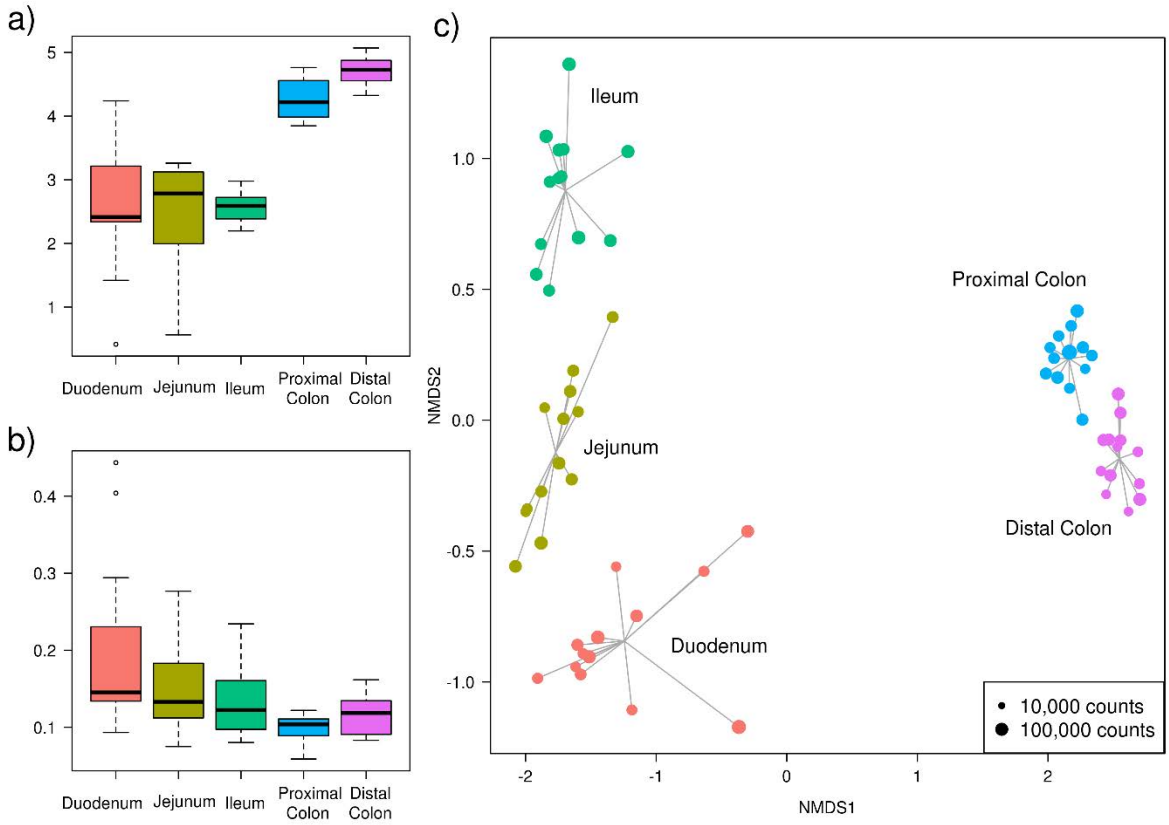


Figure 2. Descriptive plots made from the OTUs obtained in each sample. **a)** Boxplot of the Shannon α-diversity for the 13 pigs in each intestinal section. **b)** Boxplot of the Whittaker β-diversity for the 13 pigs in each intestinal section. **c)** Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities for the 65 samples of the 13 pigs in each of the 5 intestinal sections (represented by colours). The size of the dot is proportional to the total number of counts in each sample, as represented in the bottom-right rectangle.

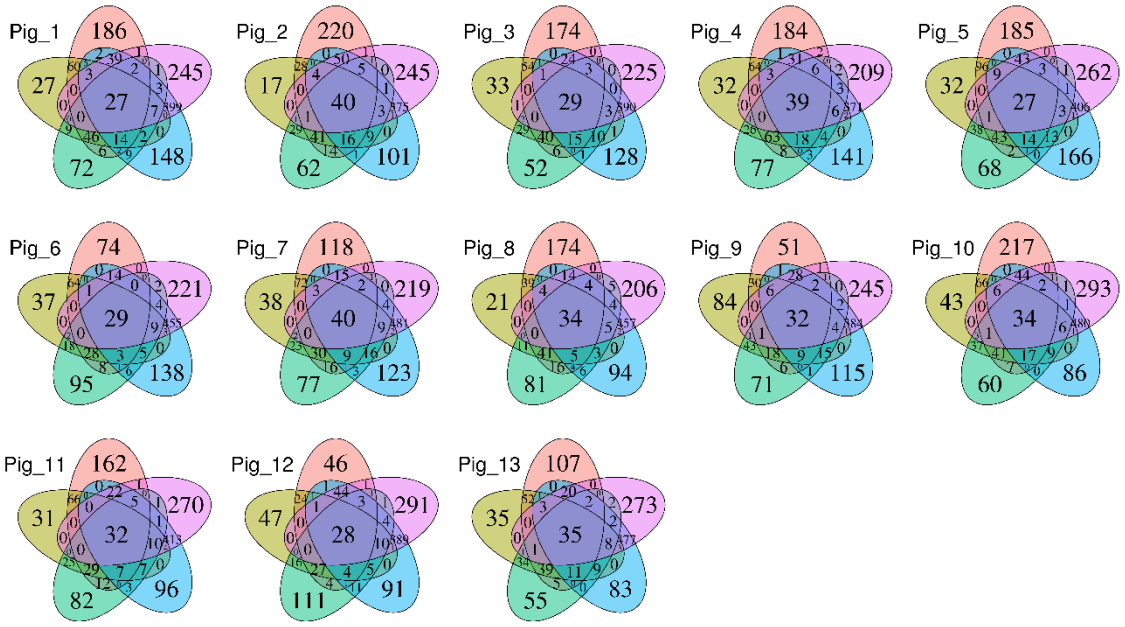


Figure 3. Five-part Venn diagram for each of the 13 subjects, showing the OTUs shared among the intestinal sections: duodenum (red), jejunum (yellow), ileum (green), proximal colon (blue), and distal colon (purple). The numbers in the diagrams represent how many OTUs were unique in the five intestinal sections or shared between sections as their areas intersect.

Microbiota interaction network reconstruction among the gut sections

To infer the interaction patterns as well as the hub genera in each section, microbial interaction networks were calculated for each gut section by using the SPCIT method as proposed by Ramayo-Caldas *et al.*[13]. For this network analysis, the correlations between genera were estimated using the relative abundances of each genus across the animals (n=13) in each of the 5 intestinal segments. In the network, every node represents a genus, and every edge connecting two nodes represents a SPCIT significant correlation: only those correlations above an absolute value of 0.65 were represented (Figure 4).

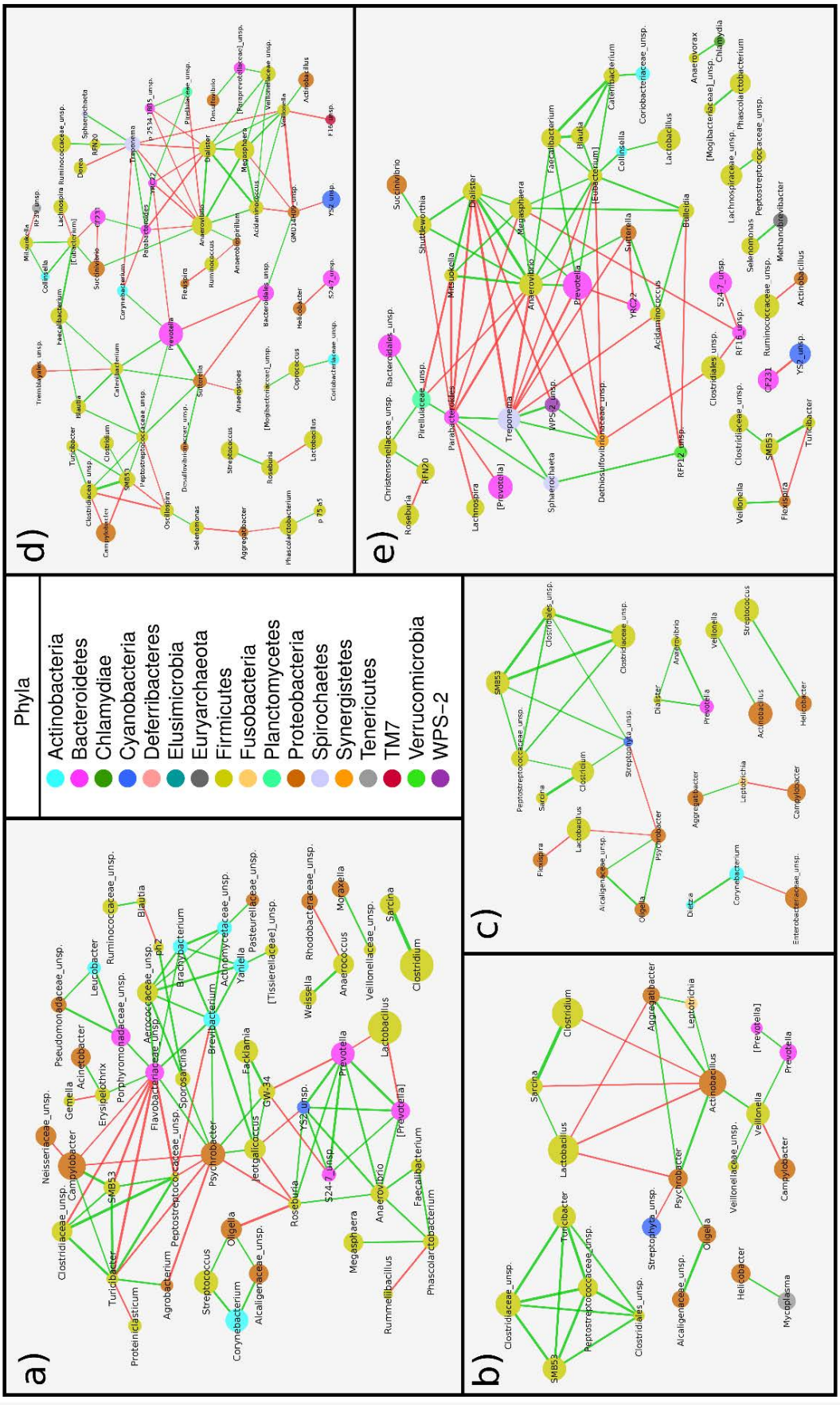


Figure 4. SparCC[55] PCIT[56] (SPCIT[13]) prokaryotic genus network for partial correlations with an absolute value above 0.65 between log-transformed genus abundances performed in the **a)** duodenum, **b)** jejunum, **c)** ileum, **d)** proximal colon, and **e)** distal colon. The width of the edge represents the degree of the correlation (wider if it is higher), and the colour shows the sign of the correlation: negative (red) and positive (green). The area of the node is proportional to the relative abundance of the prokaryotic genus. The suggested annotation for some genera is enclosed in square brackets.

In the duodenum (Figure 4a), *Psychrobacter* spp., *Jeotgalicoccus* spp. and the unspecified Flavobacteriaceae genera were part of a sub-network showing a negative correlation with another sub-network formed by *Campylobacter* spp. and *SMB53* spp. The most abundant genus in the duodenum was *Lactobacillus*, which was negatively correlated with *Prevotella* spp. The other two most abundant genera were *Clostridium* and *Sarcina*, both belonging to the Clostridiales order and showing a strong positive correlation between them, which was maintained along all the small intestine until the disappearance of *Sarcina* spp. in the hindgut.

Regarding the jejunum (Figure 4b), *Lactobacillus* spp. were associated with the two strongly linked Clostridiales genera *Clostridium* and *Sarcina*. Conversely, the *Actinobacillus* and *Psychrobacter* genera were negatively correlated to this sub-network. However, the rest of the Clostridiales order were correlated only among themselves.

In the ileum (Figure 4c), the strong correlation within the Clostridiales order, including members such as the paired *Clostridium-Sarcina* genera, was maintained. It should be noted that *Lactobacillus* spp. were still negatively correlated with *Psychrobacter* spp. Moreover, in spite of the abundance of the *Actinobacillus* and *Streptococcus* genera, they were correlated only with *Veillonella* spp. and *Helicobacter* spp., respectively.

The number of significant correlations increased in the proximal colon (Figure 4d), where the most abundant *Prevotella* spp. took the central role. *Prevotella* spp. were strongly positively correlated with *Sutterella* spp., and both of them with the Clostridiales order group. In contrast, *Campylobacter* spp. were negatively correlated with the Clostridiales order group. In addition, *Prevotella* spp. showed a negative correlation with the sub-network formed by *Treponema* spp. and *Parabacteroides* spp. This last sub-network was also in opposition to the one formed by the *Anaerovibrio*, *Dialister,* and *Megasphaera* genera.

In the last section, the distal colon (Figure 4e), *Prevotella* spp. were positively correlated with the sub-network of the *Anaerovibrio*, *Dialister,* and *Megasphaera* genera, which were negatively correlated with the *Treponema* and *Parabacteroides* sub-network, as observed in the proximal colon. *Lactobacillus*

spp. were likely to be in the same group as *Prevotella* spp., and the Clostridiales order group was not clearly correlated with the other sub-networks.

Presence/absence and differential abundance analysis of genera between consecutive sections

To compare the relative abundances of genera between the sections, we analysed each pair of consecutive regions by using metagenomeSeq[24]. After determining the presence of the genera, a differential abundance analysis was performed to filter out any genera that were absent in either of the two compared regions. The four comparisons of the two types of tests (presence/absence and differential abundance) are shown in the supplementary information found online as Supplementary Table S1 and Table S2, respectively.

Regarding the first comparison, duodenum versus jejunum, it was clearly observed that the duodenum contained more unique genera than the jejunum (38 against two). In the duodenum, there were several genera of the Actinomycetales, Bacillales, and Clostridiales orders not present in the jejunum, while *Actinomyces* spp. and *Catenibacterium* spp. were present in only the jejunum. However, there were few differentially abundant genera in this comparison (five in the duodenum and seven in the jejunum). The five most abundant bacterial genera in the duodenum were distributed in phyla as follows: one Cyanobacteria, two Firmicutes (both from the Lactobacillales order) and two Proteobacteria (belonging to the Moraxellaceae family). In contrast, the jejunum had more Firmicutes (*Turicibacter* spp. and three Clostridiales) and three Proteobacteria (*Helicobacter* spp., *Actinobacillus* spp. and an unspecified Enterobacteriaceae genus).

For the jejunum versus ileum comparison, the jejunum retained the two specific genera (*Actinomyces* and *Catenibacterium*) observed in the previous comparison, as well as two additional taxa, Bacteroidetes and *Blautia* spp., as differentially present. On the other hand, only one genus (*Flexispira* spp.) was differentially present in the ileum when compared with the jejunum. In the differential abundance study, only the Cyanobacteria phylum, which was previously more abundant in the duodenum, was shown to have a higher abundance in the jejunum than in the ileum. The eight genera that were more

abundant in the ileum include four Clostridiales and other genera such as *Actinobacillus* and *Streptococcus*.

As expected from the great distance observed in the NMDS plot between the small and large intestine samples, the comparison between the ileum and proximal colon showed the highest divergence, with a total of 64 genera being differentially present in both sections. Of these genera, only eight were differentially present in the ileum, and the rest of them (56) were in the proximal colon. The Cyanobacteria phylum was present in only the ileum when comparing it with the proximal colon, as well as other genera, such as *Dietzia*, *Facklamia* and *Sarcina*. In contrast, of the 56 genera differentially present in the proximal colon, 26 were from the Firmicutes phylum (20 from the Clostridiales order including *Butyrivibrio* spp., *Roseburia* spp. and *Ruminococcus* spp., and 6 from the Erysipelotrichales order). From the remaining 30 genera differentially present in the proximal colon, 9 belonged to the Bacteroidetes phylum, 9 were members of the Proteobacteria phylum, one was a species of Archaea from the *Methanobrevibacter* genus, and the other 11 included genera such as *Treponema* and *Chlamydia*. However, the number of differentially abundant genera between the two sections decreased to 22, with 15 genera significantly more abundant in the ileum: 8 Firmicutes (five Clostridia, including the *Veillonella* and *Clostridium* genera, and three Bacilli, *Lactobacillus* spp., *Streptococcus* spp. and *Turicibacter* spp.), 5 Proteobacteria (including *Actinobacillus* spp., *Psychrobacter* spp. and *Flexispira* spp.), and *Corynebacterium* and *Mycoplasma*. In contrast, seven genera were more abundant in the proximal colon, six of them belonging to the Clostridiales order (5 inside the Veillonellaceae family), and one from the Bacteroidales order (*Prevotella* spp.).

It is noteworthy that no genus was found to be differentially present in the proximal colon when compared with the distal colon. However, nine genera were present in only the distal colon: *Fibrobacter* spp., *Anaerovorax* spp. and two Archaea related to methane metabolism, among others. Additionally, the differential abundance study between these hindgut sections pointed out 15 genera with higher abundance in the proximal colon: one genus of the Cyanobacteria phylum, one from the Deferribacteres phylum, and seven Clostridiales genera from the Clostridiaceae and Veillonellaceae families, as well

as six genera from the Proteobacteria phylum (including *Campylobacter* spp., *Helicobacter* spp. and *Actinobacillus* spp.). In the distal colon, of the 16 differentially abundant genera in comparison to the proximal colon, 5 belonged to the Bacteroidales order, 4 were Clostridiales, and from the rest, *Treponema* spp. and a species of Archaea (*Methanobrevibacter* spp.) stood out.

Functional analysis of the gut metagenome along the intestine

Finally, PICRUSt[25] was used for metagenomic functional prediction of each of the five regions. PICRUSt utilizes 16S rRNA gene information to estimate the gene families of Archaea and bacteria that contribute to a metagenome. An NMDS analysis was performed to determine whether sample distribution depended on the predicted KEGG[26] orthologies (KOs) table (Figure 5). This plot depicts how the separation between the small and large intestines was still clearly maintained (from left to right in the plot). In addition, the proximity of the hindgut samples reveals that they are more likely to perform the same functions, which is in accordance with the similarities found in their microbiota compositions with respect to β-diversity (Figure 2b). Conversely, this NMDS plot (Figure 5), made from the predicted KOs, showed that some of the small intestine samples of one section (e.g., jejunum) were closer to samples collected from other small intestine sections (e.g., duodenum and ileum) than to samples from their own section (e.g., jejunum).

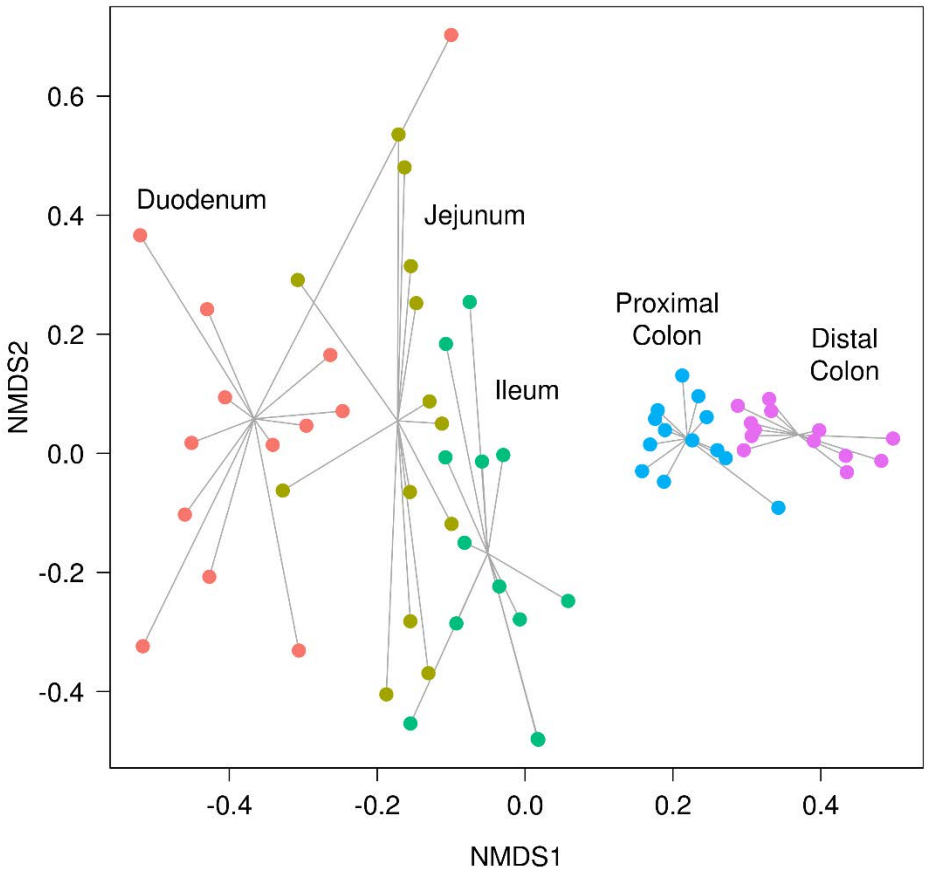


Figure 5. Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities for the metagenome (KEGG[26] orthologies (KOs) counts) predicted through PICRUSt[25] for the 65 samples from the 13 pigs in each of the 5 intestinal sections (represented by colours). In this plot, it can be seen how the predicted functions for the microbiota of the large intestine sections are more similar among individual pigs, while the predicted functions for the microbiota of the small intestine sections have more variation among individual pigs, meaning that large intestinal microbiotas are more likely to perform similar functions.

To gain a better understanding of the differentially abundant functions between the regions, the KOs were collapsed to the pathway level. Then, DESeq2[27] was utilized to compare the collapsed pathway abundances between each pair of consecutive regions; a pathway was considered more abundant in one section if its adjusted *p*-value was ≤0.01. The results are shown in the supplementary information found online as Supplementary Table S3.

In the first comparison, duodenum versus jejunum, a clear enrichment for the carotenoid and flavonoid biosynthesis pathways in the duodenum was found. In addition, the antenna proteins of the photosynthesis pathway were also more

abundant in the duodenum. Nevertheless, the most abundant pathway found in the duodenum was N-glycan biosynthesis. In contrast, the transporters and energy metabolism pathways, including fructose, mannose, amino sugar and nucleotide sugar metabolism, were more abundant in the jejunum.

In the comparison between the jejunum and ileum, pathways related to photosynthesis were still abundant in the jejunum. The most abundant pathways in the jejunum compared with the ileum were the basal transcription factors and pathways related to mineral absorption and glycolysis/gluconeogenesis and, to a lesser extent, bile acid biosynthesis and lysosomes. Conversely, the most abundant pathways in the ileum were tetracycline and polyketide sugar unit biosynthesis, both related to antibiotic synthesis. To a lesser degree, the fatty acid and lipid protein biosynthesis pathways stood out, as well as those related to amino acid metabolism.

The largest number of significant differences was found between the ileum and the proximal colon. In the ileum, pathways related to transporters and transcription factors were more abundant than in the proximal colon. The tetracycline- and lipid-related pathways were maintained in the ileum, as in the previous comparison. In addition, some degradation pathways (dioxin, ketone bodies, benzoate and xylene) were also abundant in the ileum, as well as the metabolism of pyruvate and two short chain fatty acids (SCFAs), propanoate and butanoate. Finally, one of the most relevant abundant pathways in the ileum was the phosphotransferase system, which is a bacterial method of sugar uptake. In the opposite direction, in the proximal colon, the most abundant functions were carbon fixation in photosynthetic organisms, biosynthesis of vancomycin group antibiotics, and protein digestion and absorption. Additional functions more abundant in the proximal colon were the adipocytokine and PPAR signalling pathways; sphingolipid, arachidonic acid, beta-alanine and vitamin B6 metabolism; phenylalanine, tyrosine and tryptophan biosynthesis; and, similar to in the jejunum, mineral absorption and lysosomes. Lastly, it is also worth mentioning the carbon fixation pathways in prokaryotes as well as oxidative phosphorylation and the citrate cycle (Krebs cycle).

In the last comparison between the two hindgut sections, the proximal and distal colon, the most abundant differential functions in the proximal colon were the lipopolysaccharide and its protein biosynthesis; metabolism of cofactors and vitamins, such as folate and riboflavin, was also notable, as well as glutathione and arachidonic acid metabolism. However, in the distal colon, the pathways related to bacterial cell walls were significantly more abundant (pentose and glucoronate interconversions), as well as two pathways associated with amoebiasis and bacterial antibiotic production (butirosin and neomycin biosynthesis). Nevertheless, other functions related to carbohydrate, pyruvate and methane metabolism were also identified.

The methane metabolism pathway was analysed in detail at the KO level for each comparison (see Supplementary Fig. S2 online), and the results showed that methane production was more abundant in the distal colon and acetate production was significantly greater in the hindgut.

Discussion

This study characterized the composition, distribution and potential functionality of the microbiota found in the luminal content of five sections along the digestive tract of 13 Iberian pigs at 120 days of age fed with a maize- and wheat-based diet. We confirmed the existence of extensive differences in the microbiota composition along the porcine intestine, especially between the small and large intestines, and we provide additional insights on the ecosystem structure in each section and its potential functional consequences.

In the midgut (duodenum, jejunum, and ileum) we observed a low α-diversity and a high β-diversity, while the large bowel (proximal and distal colon) had a high α-diversity and a low β-diversity. The lower α-diversity (Figure 2a) of the midgut compared with the higher α-diversity of the large bowel found in our samples was previously described in the luminal contents of 300-day-old Laiwu pigs when comparing three sections (jejunum, ileum and caecum)[18]. These differences in diversity between the small and large intestines were also found in another study of the mucosa microbiota of 28-day-old pigs[19] from the Gloucestershire Old Spot breed. In addition, the β-diversity (Figure 2b) analysis pointed out that the

differences between samples in the small intestine region were higher than those in the colon. This reduction in β-diversity when descending through the gut was also described in Laiwu pigs by Yang *et al.*[18], and it can be seen in our results, both in the NMDS (Figure 2c), where the points are closer in the hindgut, and in the smaller standard errors represented in the ten most abundant genera plot (Figure 1b). On the other hand, the greatest β-diversity was observed in the duodenum samples, with two samples very differentiated from the rest. One of these is the same outlier that appears in the α-diversity plot (Figure 2a), and thus, this individual could have a reduced diversity due to some kind of asymptomatic disease.

The higher variability observed between the midgut samples may be due to the lower number of microorganisms present in these regions. Thus, the bacterial community could potentially be less stable in the midgut than in the large bowel sections because of the continuous influx of new bacteria from food, the shorter transit time and the importance of adherence to tissue or mucus[28]. Moreover, the bacteria in the small intestine may be more susceptible to a founder effect: when early colonizers arrive, they are more prone to become established and provide the nutrients to establish a certain microbiota, but the mechanisms of this effect remain unknown[28]. Nonetheless, these pigs were raised and fed together, being exposed to similar environmental conditions; therefore, the greater differences between the small intestine samples may be due to host genetic factors, such as those related to the immune responses of the animals[31].

Consistent with the differences between intestinal sections and in accordance with other studies of the luminal[17,18] and mucosal[19] microbiota in pigs, the three most abundant phyla found along the digestive tract were Bacteroidetes, Firmicutes and Proteobacteria (Figure 1a). Nevertheless, comparing one study of the luminal microbiota with another study of the mucosal microbiota is complicated. The intestinal epithelium provides an oxygen-rich environment that could be very different from that of the luminal content[29,30], differentiating the microbiota found in these two regions[19]. Furthermore, the ratios of species abundance obtained in the two luminal content studies are difficult to compare because several variables are present in the experiments: 6-month-old Large White pigs fed with a standard diet based on maize[17] and 300-day-old Laiwu pigs

fed with a maize-soybean diet[18]. In addition, it has been shown that different ages[17], feeding[32], and DNA extraction kits[33], amongst other factors, can have an impact on the observed microbiota composition. In this sense, the breed or the genetic background of the pigs might alter the microbiota composition as well[16,34]. For example, despite the limitations of comparing two studies, Zhao *et al.*[17] described in the colon of Large White pigs a low proportion of Bacteroidetes (8.5%), while we found that the *Prevotella* genus, which belongs to the Bacteroidetes phylum, represented almost 41% in the proximal colon and 35% in the distal colon of the genera found in the Iberian pig. These two breeds are very different in lipid content; the Iberian pig is characterized by a high fat infiltration rate[20], whereas the Large White pig produces leaner meat (with a lower fat content). Hence, these differences in lipid content might be due to bacteria such as *Prevotella* spp. that can degrade the proteins and polysaccharides in the plant cell wall, producing SCFAs that can be absorbed by the host[8,35] and can modify the host lipid metabolism, increasing fat retention and adipogenesis[36,37]. Further studies are required to validate such differences in microbiota composition between breeds and their possible relationships with fat metabolism.

In our study, the Bacteroidetes and Firmicutes dominated in the colon, and the Firmicutes phylum was also the most abundant in the small intestine. The presence of the Proteobacteria phylum was increased only in the small bowel, reaching its maximum in the ileum (Figure 1a). Consequently, the presence of Firmicutes is constant along the intestine. Therefore, it is not unexpected that 13 OTUs of the *Lactobacillus* genus were found inside the core microbiota of all the samples (see Supplementary Fig. S1 online). The higher number of unique OTUs present in the duodenum was probably due to the microorganisms present in the regions from the mouth to the stomach, as well as those present in undigested food. Furthermore, possible explanations of why the microbiota uniquely present in the lower intestine does not appear in the upper intestine are the more rapid transit time in the small intestine[28] and the different environmental conditions, such as oxygen concentrations[29,30], which make the settlement and growth of certain microorganisms less likely.

Considering the correlations between genera inside each section, a tendency can be observed: the members of the Clostridiales order (*Clostridium* spp., *Sarcina*

spp. and *SMB53* spp., amongst others) were always positively associated; however, their correlation with *Campylobacter* spp. differed depending on the region, positive in the duodenum and negative in the proximal colon. These findings could be explained by the oxygen concentration along the intestine, which decreases in the last intestine sections[28]. *Campylobacter* spp. are microaerophilic microorganisms, while most of the members of the Clostridiales order are obligate anaerobes[30]. In this context, the duodenum has a higher concentration of accessible oxygen[38], and both microorganisms would take up their respective niches, not allowing the Clostridiales order to undergo excessive overgrowth, and thus, they would decrease competition with *Campylobacter* spp. for resources and space. Conversely, the reduced amount of oxygen in the colon regions favours anaerobic bacteria, such as those of the Clostridiales order, while the microaerophiles, with less oxygen available, will grow less. In this way, the Clostridiales order was also positively correlated with the sub-network formed by other anaerobic genera such as *Prevotella*[39] and *Anaerovibrio*. Other genera such as *Lactobacillus*, which are facultative anaerobes[39], can be present, as previously described, along the entire digestive tract and not be affected by oxygen competition. Finally, the opposition of *Prevotella* spp. with *Treponema* spp. found in the colon has been previously described in pigs by Ramayo-Caldas *et al.*[13]. Our guess is that these two hubs (*Prevotella* spp. and *Treponema* spp.), as well as the Clostridiales order sub-network, may compete for the degradation of dietary fibre[8,13,40]. Furthermore, the interactions between microorganisms may follow a series of universal dynamics, as has been recently proposed in the human gut microbiota[41]. However, this hypothesis needs to be validated in future studies involving larger sample sizes and adding more variables, such as age, sex, diet, and genetic background, amongst others, in order to study how the microbial interactions behave and change.

The functional prediction of the metagenome was in accordance with the microorganisms present in each intestine section and the differential abundances found in each comparison. In the first two comparisons, between the duodenum versus the jejunum and the jejunum versus the ileum, pathways related to photosynthesis were present, possibly due to the presence of the Cyanobacteria and of chloroplasts that were not yet digested inside the fodder. In the jejunum,

the main microbiome functions were focused on extracting energy from carbohydrates, such as fructose and mannose, through glycolysis/gluconeogenesis. However, in the ileum, functions were related to fatty acid and pyruvate metabolism and xylene degradation, probably because of the abundance of Clostridiales[40]. Additionally, the sugar uptake functions predicted in this region were mainly associated with the phosphotransferase system. Therefore, if the genes associated with these functions are truly expressed, these findings would confirm a spatial organization of gut microbiota functions that allows the rapid internalization and conversion of the available simple carbohydrates for microbial proliferation and maintenance[42].

SCFA production is increased in the large intestine[43]. However, in the comparison between the ileum and the proximal colon, the metabolism of two SCFAs (propanoate and butanoate) was more abundant in the ileum, probably due to the relative abundance of Clostridiales in this region (Figure 1b). Nevertheless, other Clostridiales with butyric-acid activity, such as *Butyrivibrio* spp., were found in only the proximal colon.

In the last comparison, the proximal colon showed a higher abundance of *Prevotella* spp. than the distal colon. This last gut section had more OTUs belonging to the Ruminococcaceae family, which require anaerobic conditions and a carbohydrate energy source from dietary fibre, such as cellulose or xylan[44]. In accordance with the lower availability of oxygen in the large intestine, the energy pathways related to the citric acid cycle were dominant. These pathways are characteristic of the anaerobic bacteria mentioned above, such as *Prevotella* spp., which can degrade the proteins and polysaccharides in the plant cell wall, producing SCFAs[8] that can be absorbed by the host[35].

In our study, some Archaea related to methane metabolism were found to be differentially abundant in the hindgut, where methane is predominantly produced[43]. Moreover, the methane concentration increases towards the end of the intestine and can be modified depending on dietary fibre content[45]. Thus, methane metabolism was analysed in more detail (see Supplementary Fig. S2 online). In this figure, methanogenesis was more abundant in the distal colon, as expected. Furthermore, a higher abundance of acetate production in the hindgut

than in the small intestine was found. Nonetheless, all these results must be considered carefully, because they are only predictions of the possible functions. Further studies regarding the metatranscriptome should be performed to clarify the real pathways that are present in these regions.

In summary, this study confirms that the energy pathways of the gut microbiome differ along its sections and represents, to the best of our knowledge, the first description of the gut microbiota composition along the intestine in Iberian pigs.

Methods

Ethics statement

The current study was performed according to the regulations of the Spanish Policy for Animal Protection RD53/2013, which complies with European Union Directive 2010/63/EU about the protection of animals used in experimentation. Pigs were housed in ITACyL animal facilities (Hontalbilla, Segovia, Spain), which meet local, national, and European requirements for Scientific Procedure Establishments. All experimental protocols were approved by the UCM (*Universidad Complutense de Madrid*) Ethics Committee, with reference number PROEX-007/15.

Animals and lumen content collection

A convenience sample of thirteen 120-day-old Iberian castrated male pigs from the Torbiscal line was chosen as a compromise between the detection of relevant effects and the technical, ethical, and economical limitations of increasing the sampling size. In addition, a reduced variability in their microbiota composition was expected among animals raised under controlled environmental conditions and a uniform diet. Pigs were fed *ad libitum* with a standard fodder based on maize, wheat, barley, and soybean, with 3,320 kcal of digestible energy and 15.6% of crude protein. The pigs were slaughtered at an average weight of 48.7 kg. For each pig individually, its gastrointestinal tract was removed from the abdominal cavity and dissected immediately to collect the luminal content of each of the 5 sections less than 30 minutes after the pig's death. The luminal contents of the 5 gut sections of each animal were gathered separately after isolating ~10

cm of each section with 2 disposable adjustable plastic clamps on each side as follows: duodenum, first part after the stomach's pyloric sphincter; jejunum, in the middle of the total small intestine's length; ileum, last part of the small intestine; proximal colon, first part of the large intestine after the ileocaecal valve; distal colon, last part of the large intestine just before the rectum. Afterward, an incision in the middle of each section was performed with a scalpel under aseptic conditions. Finally, disposable sterile syringes with enlarged openings were used to collect a total of 8 mL of luminal content through the incision in each section and animal. The luminal contents were transferred to cryotubes, immediately frozen in liquid nitrogen, and later stored at -80 °C until used.

DNA extraction and 16S rRNA gene sequencing

For each of the 65 samples (13 animals x five sections), the DNA of 0.2 g was extracted with the *PowerFecal*® (*MoBio*®) kit, following the manufacturer's recommendations, and DNA concentration and quality were measured with a *NanoDrop*® Spectrophotometer ND-1000. The V3-V4 region of the 16S rRNA gene was amplified with two 16S Amplicon PCR Primers (*Sigma-Aldrich*®): Forward, 5' TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG, and Reverse, 5' GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC. These two primers were designed following the *Illumina*® guide, *16S Metagenomic Sequencing Library Preparation*, based on the recommendations of Klindworth *et al.*[46]. The 65 PCR reactions were performed individually in a total volume of 25 µL using 12.5 ng of microbial DNA, 12.5 µL of 2x KAPA HiFi HotStart ReadyMix (*Kapa Biosystems*, Inc.) and 5 µL of each primer (1 µM) with the following program: 95 °C for 3 minutes, 25 cycles of three steps (95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s) and 72 °C for 5 minutes. The verification of the expected amplicon size (~550 bp) was done through agarose gel electrophoresis. Then, *AMPure*® XP beads (*Beckman Coulter*, Inc.) were used to perform PCR product clean-up. The Nextera XT Index Kit was used to attach the dual indices, and another round of PCR clean-up was done with *AMPure*® XP beads afterwards. Subsequently, the size (~630 bp) of the libraries from the indexed amplicons was validated with a DNA 1000 assay (*Agilent Technologies*, Inc) in a 2100 Bioanalyzer instrument (*Agilent Technologies*, Inc). Finally, the pooled libraries were sequenced in one

run of a *MiSeq*® (*Illumina*®) instrument in the SGB (*Servei de Genòmica i Bioinformàtica*, Cerdanyola del Vallès, Spain), using the *MiSeq*® Reagent Kit v2 (500-cycle format, paired-end (PE) reads). A mean of 126,549 sequences for each sample was obtained.

Taxonomic classification of the gut samples

The joining of the forward and reverse fastq files was performed with the *multiple_join_paired_ends.py* function in the QIIME pipeline[22] (1.9.1. version). Following the recommendations by Bokulich *et al.*[47] for raw data quality control and filtering steps, the sequences were filtered with a Phred score cut-off of 20 using the *split_libraries_fastq.py* command. Then, OTUs were identified with QIIME's subsampled open-reference OTU calling approach, as proposed by Rideout *et al.*[48], with the *pick_open_reference_otus.py* command and a subsampling percentage of 10% (s=0.1). After this step, QIIME was utilized to identify and remove chimaeras with BLAST[49]. Lastly, the final OTU dataset was obtained by filtering out singletons and OTUs representing less than 0.005% of the total counts in each section[47]. In this sense, counts are defined as the number of sequences from each sample that hit the OTU clusters described in the Greengenes 13.8 taxonomic database[23] or the "new OTU" clusters formed by QIIME.

Diversity studies and differences in the abundances of the gut microbiota

The calculation of α and β-diversities as well as the NMDS were performed in R (www.r-project.org) through the vegan package[50]. In NMDS, the dissimilarity between pairs of samples was estimated with the Bray-Curtis distance[51]. The 5-part Venn diagrams were represented using the *draw.quintuple.venn* function of the VennDiagram R package[52].

The OTU information table was merged at the genus level for each of the 65 samples through their available taxonomic information with the *tax_glom* method inside the phyloseq R package[53]. Then, the genera presence/absence analysis between each section and the next one was carried out with the metagenomeSeq R package[24] using its *fitPA* function after filtering out the genera that were not present in either of the two compared sections. Thus, these genera, which were determined as not present in one of the two sections with an adjusted *p*-value

≤0.01 cut-off, were not considered for the genus-level differential abundance analysis. This second analysis was also performed with metagenomeSeq utilizing its *fitZig* function to create a model where the animal was included as a co-factor and only those genera with an adjusted *p*-value ≤0.01 cut-off were kept. Adjusted *p*-values for the *fitPA* and *fitZig* results were calculated in both cases through the false discovery rate (FDR) method[54].

Metagenome prediction and functional differences amongst gut sections

The metagenome KEGG[26] orthologies (KOs) of the 65 samples were predicted with the PICRUSt software[25], removing the OTUs that were not present in the Greengenes 13.5 database[23]. Then, the Bray-Curtis distance[51] was used to measure the dissimilarity between pairs of samples to create an NMDS plot of the KOs with the vegan package[50]. After this step, the KOs were collapsed to the pathway level (KEGG level 3) with the *categorize_by_function.py* script of PICRUSt. The differences in abundance of these collapsed pathways were identified by the DESeq2 R package[27] using a model where the animal was included as a co-factor. One pathway was considered more abundant in one section than the other when its adjusted *p*-value calculated through the FDR method[54] was ≤0.01.

Network prediction with the SPCIT approach

To infer the interaction patterns as well as the hub genera in each section, the network was calculated using the SPCIT method, as proposed by Ramayo-Caldas *et al*.[13]. First, to avoid the errors caused by the small number of samples in each section, the genera were split by section and filtered out if they represented less than 0.01% of the total section counts or they were not present (equals 0) in seven or more of the 13 samples. Then, due to violation of the sparsity assumption, the Sparse Correlations for Compositional data software (SparCC)[55] advised application of the central log ratio transformation to the genus abundances in order to calculate correlations among the genera. To extract significant correlations, a strategy based on partial correlation and information theory was applied through the Partial Correlation coefficient with Information Theory (PCIT) algorithm[56]. The Cytoscape software[57] was used to represent the network of partial correlations between the log-transformed genus abundances

for each intestinal section. In the networks of the five intestinal sections, every node represents a genus, and every edge connecting two nodes represents a significant correlation. Following the recommendations by Ramayo-Caldas *et al.*[13], from the pairwise correlation matrix obtained with SparCC in each section, median + 2 * SD was calculated as the cut-off. For clarity of presentation, only correlations above an absolute value of 0.65 were represented, as this value was the median cut-off of the five sections.

Data Availability

The raw sequencing data from this study were deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP136308.

Acknowledgements

This work was funded by the *Ministerio de Economía y Competitividad* (MINECO) and the *Fondo Europeo de Desarrollo Regional* (FEDER) projects AGL2014-56369-C2-2-R and AGL2017-82641-R. D. Crespo-Piazuelo was funded by a "*Formació i Contractació de Personal Investigador Novell*" (FI-DGR) PhD grant from the *Generalitat de Catalunya* (ECO/1788/2014) and by the PiGutNet COST Action (www.pigutnet.eu) for a Short Term Scientific Mission at the GABI laboratory (INRA, France) under the supervision of J. Estellé. M. Revilla was also funded by a FI-DGR (ECO/1639/2013). L. Criado-Mesas was funded with an FPI grant from the AGL2014-56369-C2 project. M. Ballester was financially supported by a "*Ramón y Cajal*" contract (RYC-2013-12573) from the Spanish Ministry of Economy and Competitiveness. The animal material was obtained in the "*Centro de Pruebas de Porcino del ITACyL*" in Hontalbilla (Segovia, Spain) provided by the MEDGAN-CM S2013/ABI-2913 project. We acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the "*Severo Ochoa Programme for Centres of Excellence in R&D*" 2016-2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics and the CERCA Programme / *Generalitat de Catalunya*.

Authors' Contributions

JMF and AIF conceived and designed the experiments; JMF was the principal investigator of the project; this work is part of the PhD thesis of DCP, co-supervised by MB and JMF; DCP, JE, MR, CO and JMF collected samples; DCP, LCM, and MB performed the microbial DNA extraction; DCP, JE, and YRC analysed the data; DCP, MB, and JMF wrote the paper. All authors read and approved the final manuscript.

Additional information

Competing interests

The authors declare no competing interests.

References

1. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307,** 1915–20 (2005).

2. Yang, X., Xie, L., Li, Y. & Wei, C. More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *PLoS One* **4,** e6074 (2009).

3. Xiao, L. *et al.* A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* **1,** 16161 (2016).

4. Wang, M. & Donovan, S. M. Human microbiota-associated swine: current progress and future opportunities. *ILAR J.* **56,** 63–73 (2015).

5. Nguyen, T. L. A., Vieira-Silva, S., Liston, A. & Raes, J. How informative is the mouse for human gut microbiota research? *Dis. Model. Mech.* **8,** 1–16 (2015).

6. Kuczynski, J. *et al.* Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* **13,** 47–58 (2011).

7. Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. Analysis

of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* **469,** 967–77 (2016).

8. Liu, H., Ivarsson, E., Dicksved, J., Lundh, T. & Lindberg, J. E. Inclusion of Chicory (Cichorium intybus L.) in pigs' diets affects the intestinal microenvironment and the gut microbiota. *Appl. Environ. Microbiol.* **78,** 4102–4109 (2012).

9. Ivarsson, E., Roos, S., Liu, H. Y. & Lindberg, J. E. Fermentable non-starch polysaccharides increases the abundance of Bacteroides-Prevotella-Porphyromonas in ileal microbial community of growing pigs. *Animal* **8,** 1777–87 (2014).

10. Mach, N. *et al.* Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ. Microbiol. Rep.* **7,** 554–569 (2015).

11. Slifierz, M. J., Friendship, R. M. & Weese, J. S. Longitudinal study of the early-life fecal and nasal microbiotas of the domestic pig. *BMC Microbiol.* **15,** 184 (2015).

12. Kraler, M., Ghanbari, M., Domig, K. J., Schedle, K. & Kneifel, W. The intestinal microbiota of piglets fed with wheat bran variants as characterised by 16S rRNA next-generation amplicon sequencing. *Arch. Anim. Nutr.* **70,** 173–89 (2016).

13. Ramayo-Caldas, Y. *et al.* Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J.* **10,** 2973–2977 (2016).

14. He, M. *et al.* Evaluating the Contribution of Gut Microbiota to the Variation of Porcine Fatness with the Cecum and Fecal Samples. *Front. Microbiol.* **07,** 2108 (2016).

15. McCormack, U. M. *et al.* Exploring a Possible Link between the Intestinal Microbiota and Feed Efficiency in Pigs. *Appl. Environ. Microbiol.* **83,** e00380-17 (2017).

16. Camarinha-Silva, A. *et al.* Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs. *Genetics*

206, 1637–1644 (2017).

17. Zhao, W. *et al.* The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments. *PLoS One* **10,** 1–13 (2015).

18. Yang, H. *et al.* Uncovering the composition of microbial community structure and metagenomics among three gut locations in pigs with distinct fatness. *Sci. Rep.* **6,** 27427 (2016).

19. Kelly, J. *et al.* Composition and diversity of mucosa-associated microbiota along the entire length of the pig gastrointestinal tract; dietary influences. *Environ. Microbiol.* **19,** 1425–1438 (2017).

20. Serra, X. *et al.* A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs. *Livest. Prod. Sci.* **56,** 215–223 (1998).

21. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 15718–23 (2004).

22. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7,** 335–6 (2010).

23. DeSantis, T. Z. *et al.* Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072 (2006).

24. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10,** 1200–2 (2013).

25. Langille, M. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31,** 814–21 (2013).

26. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

27. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550

(2014).

28. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14,** 20–32 (2016).

29. Marteyn, B., Scorza, F. B., Sansonetti, P. J. & Tang, C. Breathing life into pathogens: the influence of oxygen on bacterial virulence and host responses in the gastrointestinal tract. *Cell. Microbiol.* **13,** 171–6 (2011).

30. Albenberg, L. *et al.* Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology* **147,** 1055–1063.e8 (2014).

31. Fagarasan, S. *et al.* Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* **298,** 1424–7 (2002).

32. Burrough, E. R., Arruda, B. L., Patience, J. F. & Plummer, P. J. Alterations in the Colonic Microbiota of Pigs Associated with Feeding Distillers Dried Grains with Solubles. *PLoS One* **10,** e0141337 (2015).

33. Wagner Mackenzie, B., Waite, D. W. & Taylor, M. W. Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences. *Front. Microbiol.* **6,** 130 (2015).

34. Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl. Acad. Sci.* **107,** 18933–18938 (2010).

35. Brüssow, H. & Parkinson, S. J. You are what you eat. *Nat. Biotechnol.* **32,** 243–245 (2014).

36. Jørgensen, H., Larsen, T., Zhao, X. Q. & Eggum, B. O. The energy value of short-chain fatty acids infused into the caecum of pigs. *Br. J. Nutr.* **77,** 745–56 (1997).

37. Canfora, E. E., Jocken, J. W. & Blaak, E. E. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat. Rev. Endocrinol.* **11,** 577–91 (2015).

38. He, G. *et al.* Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with

spatial and spectral EPR imaging. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 4586–91 (1999).

39. Bergey, D. *Bergey's manual of systematic bacteriology.* (New York: Springer, 2001).

40. Niu, Q. *et al.* Dynamic distribution of the gut microbiota and the relationship with apparent crude fiber digestibility and growth stages in pigs. *Sci. Rep.* **5,** 9938 (2015).

41. Bashan, A. *et al.* Universality of human microbial dynamics. *Nature* **534,** 259–62 (2016).

42. Zoetendal, E. G. *et al.* The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.* **6,** 1415–26 (2012).

43. Robinson, J. A., Smolenski, W. J., Ogilvie, M. L. & Peters, J. P. In vitro total-gas, CH4, H2, volatile fatty acid, and lactate kinetics studies on luminal contents from the small intestine, cecum, and colon of the pig. *Appl. Environ. Microbiol.* **55,** 2460–7 (1989).

44. Bryant, M. P. Bacterial species of the rumen. *Bacteriol. Rev.* **23,** 125–53 (1959).

45. Jensen, B. B. & Jørgensen, H. Effect of dietary fiber on microbial activity and microbial gas production in various regions of the gastrointestinal tract of pigs. *Appl. Environ. Microbiol.* **60,** 1897–904 (1994).

46. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41,** 1–11 (2013).

47. Bokulich, N. A. *et al.* Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* **10,** 57–9 (2013).

48. Rideout, J. R. *et al.* Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* **2,** e545 (2014).

49. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic

local alignment search tool. *J. Mol. Biol.* **215,** 403–10 (1990).

50. Oksanen, J. *et al.* vegan: Community Ecology Package. (2016).

51. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27,** 325–349 (1957).

52. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12,** 35 (2011).

53. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8,** e61217 (2013).

54. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57,** 289–300 (1995).

55. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8,** e1002687 (2012).

56. Reverter, A. & Chan, E. K. F. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* **24,** 2491–2497 (2008).

57. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–504 (2003).

Supplementary Tables

Supplementary Table S1. This table contains the results of the presence/absence analysis at the genus level performed with metagenomeSeq[24]. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which genus is uniquely present in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon.

Supplementary Table S2. This table contains the results of the differential abundance analysis at the genus level performed with metagenomeSeq[24]. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which genus is more abundant in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and, purple, distal colon.

Supplementary Table S3. This table contains the results of the differential abundance analysis performed with DESeq2[27] for the KEGG[26] orthologies (KOs) predicted with PICRUSt[25] at the pathway level. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which pathway is more abundant in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon.

Supplementary Figures

Supplementary Figure S1. Five-part Venn diagram performed for the OTUs shared among sections when combining the datasets from all subjects: duodenum (red), jejunum (yellow), ileum (green), proximal colon (blue), and distal colon (purple).

Supplementary Figure S2. DESeq2[27] results below a padj≤0.01 cut-off for the four comparisons between each pair of consecutive sections of the KEGG[26] orthologies (KOs) predicted by PICRUSt[25] represented over the KEGG[26] methane metabolism pathway (map00680): **I.** duodenum vs jejunum; **II.** jejunum vs ileum; **III.** ileum vs proximal colon; **IV.** proximal colon vs distal colon. The colour shows which KO was more abundant in that section when comparing the two sections: white, non-significant; red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon. Ambiguous KOs were coloured grey. For clarity of presentation, the methane metabolism pathway was divided into two parts: **a)** The red rectangle shows how the production of methane was more abundant in the distal colon than in the rest of the comparisons. **b)** The two red rectangles represent how the production of acetate was more abundant in the proximal colon than in the ileum and more abundant in the distal colon than in the proximal colon.

Paper IV

Association between the pig genome and its gut microbiota composition

Crespo-Piazuelo, D., Migura-Garcia, L., Estellé, J., Criado-Mesas, L., Revilla, M., Castelló, A., Muñoz, M., García-Casco, J. M., Fernández, A. I., Ballester, M., and Folch, J. M.

The ISME Journal (*in revision*)

Association between the pig genome and its gut microbiota composition

Running title: Association of genome with gut microbiota in pigs

Daniel Crespo-Piazuelo[1,2], Lourdes Migura-Garcia[3], Jordi Estellé[4], Lourdes Criado-Mesas[1], Manuel Revilla[1,2], Anna Castelló[1,2], María Muñoz[5,6], Juan M García-Casco[5,6], Ana I Fernández[5], Maria Ballester[3], and Josep M Folch[1,2]

[1]Plant and Animal Genomics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, Bellaterra, Spain

[2]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain

[3]Departament de Genètica i Millora Animal, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Caldes de Montbui, Spain

[4]Génétique Animale et Biologie Intégrative (GABI), Institut National de la Recherche Agronomique (INRA), AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

[5]Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain

[6]Centro I+D en Cerdo Ibérico INIA-Zafra, Zafra, Spain

Address correspondence to Daniel Crespo-Piazuelo: Edifici CRAG, Campus UAB, 08193 Cerdanyola del Vallès, Barcelona, Spain; Telephone number: +34 935 636 600; E-mail: daniel.crespo@cragenomica.es

The authors declare that they have no competing interests.

Abstract

The gut microbiota has been evolving with its host along the time creating a symbiotic relationship. In this study, we assess the role of the host genome in the modulation of the microbiota composition in pigs. Gut microbiota compositions were estimated through sequencing the V3-V4 region of the 16S rRNA gene from rectal contents of 285 pigs. A total of 1 257 operational taxonomic units were obtained and grouped in 18 phyla and 101 genera. *Firmicutes* (45.36%) and *Bacteroidetes* (37.47%) were the two major phyla obtained, whereas at genus level *Prevotella* (7.03%) and *Treponema* (6.29%) were the most abundant. Pigs were also genotyped with a high-throughput method for 45 508 single nucleotide polymorphisms that covered the entire pig genome. Subsequently, genome-wide association studies were made among the genotypes of these pigs and their gut microbiota composition. A total of 52 single-nucleotide polymorphisms distributed in 17 regions along the pig genome were associated with the relative abundance of six genera; *Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53*, and *Streptococcus*. Our results suggest 38 candidate genes that may be modulating the microbiota composition and manifest the association between host genome and gut microbiota in pigs.

Keywords

16S rRNA gene, pig gut, microbiota composition, pig genome, GWAS

Introduction

The digestive tract of animals has been evolving along the time with symbiotic microorganisms. These microbes, mostly bacteria, have adapted to thrive in such conditions forming complex and vital interactions among them and their host [1, 2]. The ecological community of these microorganisms is called microbiome, and the interactions with the host can be commensal, pathogenic or mutualistic [3]. In this scenario, mutualistic gut microbiota provides the host with beneficial functions that the host cannot perform, such as digesting complex polysaccharides, producing vitamins, and preventing colonization by pathogens [2, 4]. Likewise, commensal gut populations modulate hosts' immune responses which can modify the microbiota composition in order to maintain gut homeostasis [4]. Therefore, apart from the host genetics, the complexity of the interactions increases taking into account factors such as age, diet, environment, disease, or maternal seeding which are known to influence gut microbial communities [5].

The intestinal epithelium acts as a barrier, protecting deeper tissues from bacterial entry [2]. Supporting this defence system, the gut epithelial surface is coated with a mucous layer formed by mucin glycoproteins [6, 7]. While the small intestine has only one layer which is permeable to bacteria [6], the mucous layer of the colon is structured in two parts: a dense inner layer firmly attached to the gut epithelium that minimizes bacterial-epithelial cell contact, and a loose outer layer that can be broken down by commensal bacteria [7]. In this outer mucous layer, the metabolites produced by these bacteria interact with the host stimulating the innate and adaptive immune responses [2]. For instance, host innate immunity can select for a species-specific microbiota using microbicidal proteins [8]. However, the host also has mechanisms to tolerate the metabolites from non-pathogenic bacteria [2], just as certain bacteria trigger the host immune system for self-benefit [9].

In these recent years, high-throughput sequencing technologies have greatly improved the study of bacterial populations without performing microbial cultures. The microbial 16S rRNA gene sequencing is commonly used to estimate the microbiota composition, while the shotgun sequencing of DNA fragments isolated after shearing faecal or other samples is used for the metagenome (all the

microbial collective genomes) characterization [10]. Recently, whole-metagenome sequencing has been used to obtain the reference gene catalogue of the pig gut microbiome [11]. This study revealed that the reference catalogue of the porcine gut microbiome shared more non-redundant genes between human and pig than human and mouse [11], suggesting pig as a better animal model than mouse because of their similarity with humans. Both species are omnivores and have monogastric digestive tracts which are analogous in anatomy, immunology and physiology [12].

The heritability of the microbial genera composition of the pig gut has been reported to range from low to high values [13, 14]. Accordingly, host genetics has been suggested as an important factor in the determination of gut microbial composition [15]. However, there are limited studies measuring the contribution of inter-individual variability modulating the bacterial communities and the effect of host polymorphisms on the establishment of the microbiota [16]. In this context, genome-wide association studies (GWAS), which have been widely used to analyse a plethora of complex traits, are now being used to study the link between the host and its microbiota composition [15, 16]. With this approach, Blekhman *et al*. [17] were the first to describe in humans the relationship between the abundance of *Bifidobacterium* and the single-nucleotide polymorphisms (SNPs) close to the lactase gene. In this case, lactase non-persistent recessive individuals who drink milk cannot break down lactose and thus, *Bifidobacterium* thrives using this available sugar [18].

Conversely, host genetics appeared to have a minor impact in the microbiota compared with age, diet or the environment [19]. It is not surprising, since conditions are difficult to standardize between individuals. In this regard, production pigs represent a perfect model to measure the effect of host genetics in shaping the microbiota due to their similar diet and environmental factors during their whole rearing cycle, but the relationship between the pig genome and its gut microbiota composition has not yet been fully described [20].

The objective of this study was to identify genomic regions that influence the gut microbiota composition through host-microbiota associations in pigs. For this purpose, the 16S rRNA gene was sequenced from rectal contents of 288 pigs genotyped with a high-throughput method.

Materials and Methods

Ethics approval

All animal manipulations were performed according to the regulations of the Spanish Policy for Animal Protection RD53/2013, which meets the European Union Directive 2010/63/EU about the protection of animals used in experimentation. Pigs were slaughter in a commercial abattoir following national and institutional guidelines for Good Experimental Practices.

Animal material

A total of 288 healthy commercial F1 crossbred pigs (Duroc × Iberian) were used in this study. All animals were maintained in the same farm under intensive conditions and feeding was *ad libitum* with a barley- and wheat-based commercial diet. Pigs with an average weight of 138.8 kg (SD=11.46 kg) were slaughtered in a commercial abattoir in four distinct days. Samples of rectal content and *Longissimus dorsi* muscle were snap-frozen in liquid nitrogen and later stored at –80 °C.

Microbial DNA extraction and 16S rRNA gene sequencing

For each one of the 288 samples, the DNA of 0.2 g of rectal content was extracted with PowerFecal kit (MoBio Laboratories, Carlsbad, CA, USA) following the manufacturer's recommendations. DNA purity and concentration were measured through a ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The amplification of the V3-V4 region of the 16S rRNA gene was performed following the recommendations of the *16S Metagenomic Sequencing Library Preparation* guide (Illumina, San Diego, CA, USA). Full description of primer sequences and methods used can be accessed at Supplementary Information S1. All the 288 amplicon pooled libraries were sequenced in three runs of a MiSeq (Illumina, San Diego, CA, USA) instrument in the Sequencing Service of the FISABIO (*Fundació per al Foment de la Investigació Sanitària i Biomèdica de la Comunitat Valenciana*, Valencia, Spain) using the MiSeq Reagent Kit v3 (600-cycle format, 2×300 bp paired-end reads) and a mean of 104 115 reads for each sample was obtained (17.991 Gb in total).

Taxonomy classification and diversity studies of the gut samples

Bioinformatics analysis were performed in QIIME v.1.9.1 [21] by using the QIIME's subsampled open-reference operational taxonomic unit (OTU) calling

approach and following the recommendations of Rideout *et al*. [22]. In brief, the *join_paired_ends.py* function in QIIME was used to merge the forward and reverse reads contained in the fastq files of the 288 samples. The quality control and the filtering process was made pursuant to the considerations provided by Bokulich *et al*. [23]. Therefore, the *split_libraries_fastq.py* command was used to demultiplex and filter (at Phred ≥ Q20) the fastq sequence data. After this step, OTUs were identified by using the *pick_open_reference_otus.py* function with a subsampled percentage of 10% (s=0.1). Subsequently, chimera detection was carried on in QIIME with BLAST [24] and OTUs were taxonomically annotated employing the Greengenes 13.8 database [25]. At this point, three samples did not satisfy the quality filters and were discarded. Thus, for the remaining 285 samples, a dataset containing 1 269 OTUs was obtained after filtering out singletons and OTUs representing less than 0.005% of the total number of annotated reads [23]. From this dataset, 12 *de novo* OTUs were not classified in any taxonomic rank and they were discarded. Finally, 1 257 OTUs in the 285 samples were considered for further analysis.

The 1 257 OTUs were grouped in 18 phyla and 101 genera through the *tax_glom* method within the phyloseq package [26] in R (www.r-project.org). Besides, genera that belonged to a higher taxonomy rank but lacked the genus information were merged and marked as unspecified (g__unsp).

The analyses of α and β-diversities in the 285 samples were carried on with the vegan R package [27], and the non-metric multidimensional scaling (NMDS) plot was performed using phyloseq [26] and ggplot2 [28]. For the α-diversity study, the Shannon index was employed, whereas the β-diversity study was represented using the Whittaker index. Additionally, the dissimilarity between pairs of samples was estimated with the NMDS method using the Bray-Curtis dissimilarity [29].

Host DNA extraction and SNP genotyping

Pig genomic DNA was extracted from the *Longissimus dorsi* muscle of all the 288 samples using the standard phenol-chloroform method [30]. The DNA concentration and purity was measured with a ND-1000 spectrophotometer (NanoDrop) afterwards.

A total of 288 pigs were genotyped with the GeneSeek Genomic Profiler (GGP) Porcine HD v1 (70K) array (Illumina, San Diego, CA, USA) using the Infinium HD Assay Ultra protocol (Illumina). Genotypes were obtained with the GenomeStudio software (2011.1 version, Illumina) and filtered with the PLINK software [31] (1.90b5 version). Further analyses were conducted using only SNPs that mapped in the *Sscrofa11.1* assembly, with a minor allele frequency (MAF) > 5% and missing genotypes < 5%, retaining a total of 45 508 SNPs.

GWAS analysis

For the 285 samples, GWAS between the microbiota composition at genus level and the 45 508 genotyped SNPs were made. Annotated reads of genera were normalized in relative percentages within each one of the 285 samples. To avoid errors caused by low abundant genera, GWAS were performed only in genera that comprised more than the 0.5% of the total annotated reads and were present in more than the 90% of the samples. In addition, genera marked as unspecified were excluded from the GWAS analysis. Therefore, GWAS were performed in 18 of the 101 genera found.

For the GWAS analysis, the following univariate linear mixed model was applied using the GEMMA software [32] (0.96 version):

$y_{ijkl} = Sex_i + Batch_j + u_k + \lambda_k a_l + e_{ijkl}$,

where y_{ijkl} indicates the vector of phenotypic observations in the k^{th} individual; sex (two categories) and batch (4 categories) are fixed effects; u_k is the infinitesimal genetic effect considered as random and distributed as $N(0, A\sigma_u)$, where A is the numerator of the kinship matrix; λ_k is a -1, 0, +1 indicator variable depending on the k^{th} individual genotype for the l^{th} SNP; a_l represents the additive effect associated with the l^{th} SNP; and e_{ijkl} is the residual.

The false discovery rate (FDR) method developed by Benjamini and Hochberg [33] was applied for multiple test correction using the *p.adjust* function incorporated in R. The cut-off for considering a SNP as significant was set at FDR≤0.1. Two significant SNPs were grouped inside the same interval if the distance between them was less than 2 Mb.

Gene annotation and functional prediction

The associated regions in the pig genome were annotated at 1 Mb on each side of the previously defined intervals. The genes contained in these regions were

extracted using the BioMart tool [34] from the Ensembl project (www.ensembl.org; release 92) using the *Sscrofa11.1* reference assembly. In addition, the functional consequences of the significant SNPs were predicted through the Variant Effect Predictor tool [35] from the Ensembl project (release 92).

Results and Discussion

Microbiota composition and diversity

A mean of 104 115 reads per sample were obtained with a MiSeq (Illumina) after sequencing the V3-V4 region of the 16S rRNA gene from rectal contents of 288 pigs. A total of 1 257 OTUs which were grouped in 18 phyla and 101 genera were found in the 285 samples that fulfilled the quality criteria. At phylum level, *Firmicutes* (45.36%) and *Bacteroidetes* (37.47%) were the more abundant (Fig. 1 and Supplementary Table S1). In accordance with the literature [36, 37], *Firmicutes* and *Bacteroidetes* are usually the most dominant phyla found in colon and faeces of pigs. The most abundant genera, not marked as unspecified, were *Prevotella* (7.03%) and *Treponema* (6.29%) (Supplementary Table S2). Accordingly, *Prevotella* spp. are frequently found as one of the most abundant genus in the lower intestine and faeces [36, 37]. However, comparisons between different studies should be made with caution, since differences in microbiota composition are conditional on the different sets of primers used in the analysis, breeds (host genetic background), age of the animals at sampling time, and environmental factors such as dietary composition [14, 38].

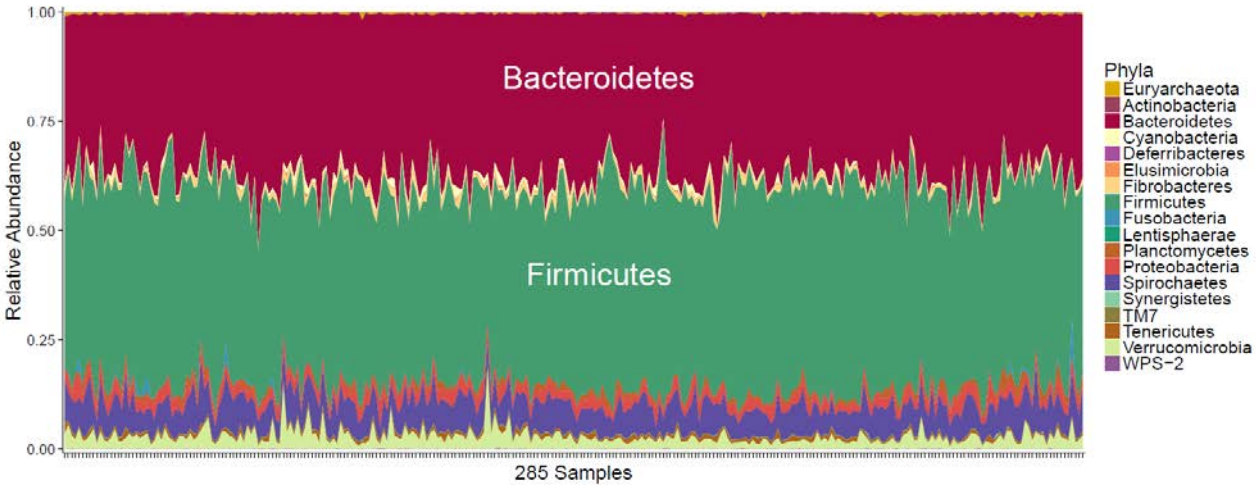


Fig. 1. Stacked area plot of OTUs grouped by phyla for the 285 pig rectal samples.

To obtain a measure for the number of different OTUs and their relative abundance within each of the 285 samples, the community α-diversity was calculated through the Shannon index (Fig. 2a). The mean of the α-diversity was 5.58, ranging from 5.00 to 5.96. It is not surprising, since the distal part of the pig gut usually has a higher α-diversity than the rest of the intestine [37]. In addition, the β-diversity was used to measure the differences between samples through the Whittaker index (Fig. 2b) obtaining a mean distance to the centroid of 0.10. Lastly, a NMDS plot was performed to observe the dissimilarities between samples employing Bray-Curtis dissimilarity (Fig. 2c). The β-diversity was closer to 0, indicating that the global microbiota composition was quite similar along the 285 samples. Furthermore, this low β-diversity was expected, since the pigs from our study have been subjected to the same diet and environmental factors during their whole rearing cycle and this uniformity is reinforced by the absence of clustering in the NMDS plot. This way, overall diversity results reinforce the appropriateness of the model to measure the effect of host genetics in shaping the microbiota.

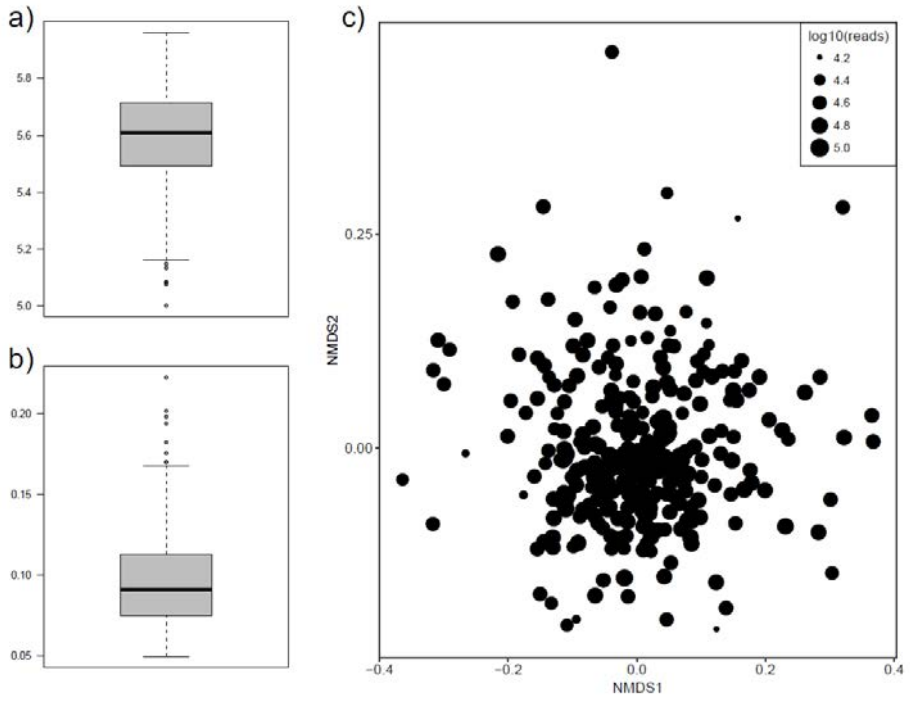


Fig. 2. Plots showing the diversities and dissimilarities measured using the 1,257 OTUs found in rectal contents of 285 pigs. a) Boxplot of the Shannon α-diversity. b) Boxplot of the Whittaker β-diversity calculated through the Bray-Curtis dissimilarity. c) Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities. The size of the dot is proportional to the total number of annotated reads in each sample.

GWAS results

GWAS were performed using 45 508 SNPs genotyped in 285 animals and the relative abundance of 18 genera; *Akkermansia*, *Bacteroides*, *CF231*, *Coprococcus*, *Fibrobacter*, *Lactobacillus*, *Oscillospira*, *Parabacteroides*, *Paraprevotellaceae Prevotella*, *Phascolarctobacterium*, *Prevotella*, *RFN20*, *Ruminococcus*, *SMB53*, *Sphaerochaeta*, *Streptococcus*, *Treponema* and *YRC22*. A total of 52 significant SNPs were distributed in 17 regions along the following *Sus scrofa* chromosomes (SSC): SSC3, SSC4, SSC6, SSC7, SSC8, SSC9, SSC10, SSC11, SSC13, SSC14, SSC15, SSC18 and SSCX (Supplementary Table S3). Significant association signals (FDR ≤ 0.1) were found in six out of the 18 GWAS for the following genera: *Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53* and *Streptococcus* (Fig. 3 and Table 1). No shared associated regions were found for the abundances of these six genera, albeit some of them belong to the same phyla. *CF231* and *Prevotella* are genera of the *Bacteroidetes* phylum. Within the *Firmicutes* phylum, *Phascolarctobacterium* and *SMB53* are members of the *Clostridiales* order, and *Streptococcus* is a member of the *Lactobacillales* order. Hence, our results suggest an association between chromosomal regions along the pig genome and abundance of certain bacteria genera. In the following sections, the candidate genes mapped in the genomic regions associated with the genus relative abundance of *Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53* and *Streptococcus* are discussed in detail. The list of candidate genes is summarized in Table 1.

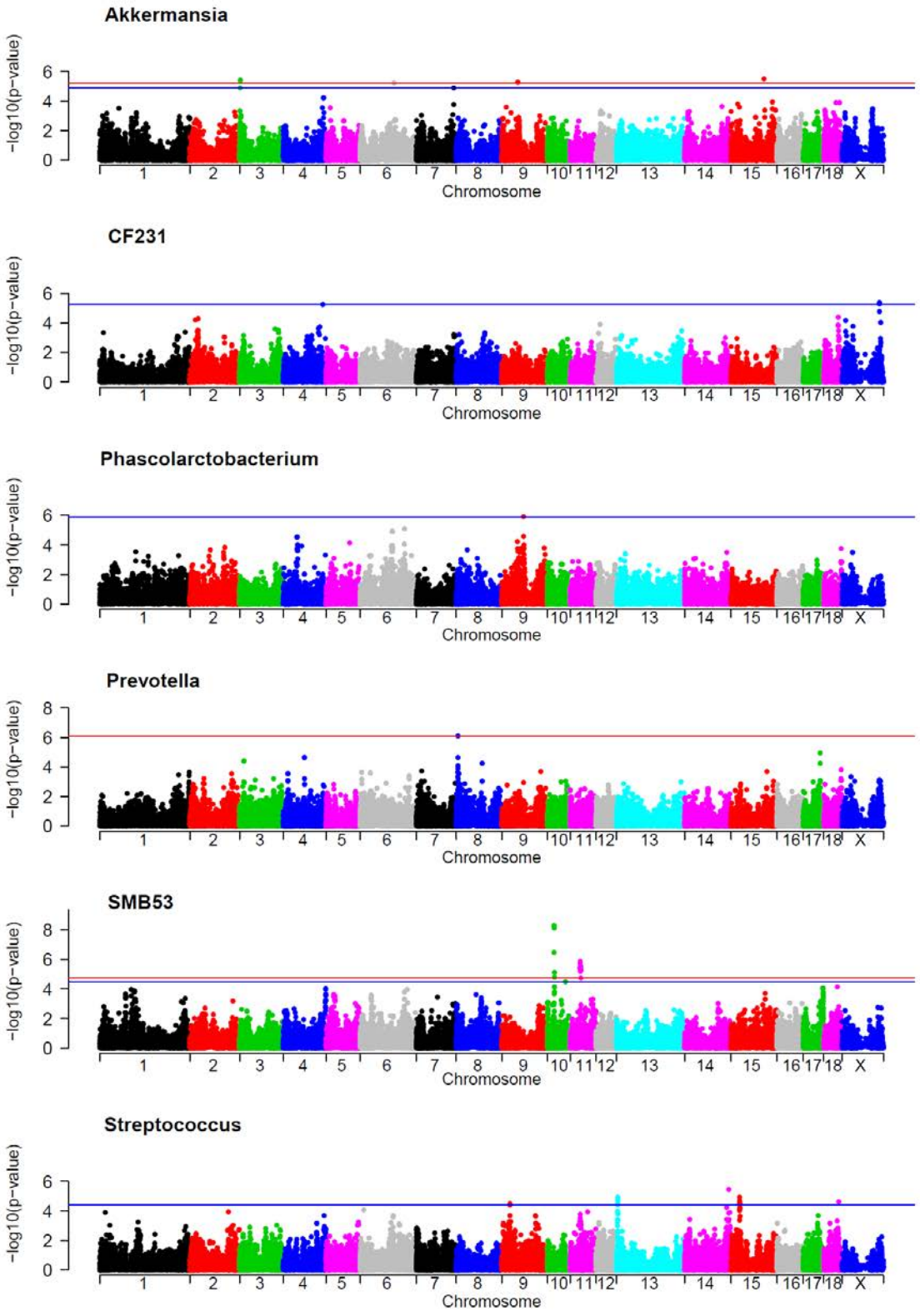


Fig. 3. GWAS plot for the relative abundance of the following genera: *Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53*, and *Streptococcus*. The red lines indicate those SNPs that are below the genome-wide significance threshold (FDR ≤ 0.05), while the blue lines indicate those SNPs that are below genome-wide significance threshold (FDR ≤ 0.1).

Table 1. Significant genomic regions in the pig genome associated with the relative composition of genera and the candidate genes found within.

Region	Genus	Chr.[a]	Position in Mb Start - End	No. SNPs[b]	Most significant SNP	Effect (%)	*p*-value	FDR	Candidate genes
A1	*Akkermansia*	3	0.89 - 3.04	3	rs81335357; rs81246645	7.72	4.04×10^{-6}	4.58×10^{-2}	*CARD11*; *CHST12*
A2	*Akkermansia*	6	101.46 - 103.46	1	rs81390429	25.85	6.06×10^{-6}	4.58×10^{-2}	*TGIF1*
A3	*Akkermansia*	7	112.61 - 114.61	1	rs325604118	6.56	1.33×10^{-5}	7.55×10^{-2}	*LGMN*; *CHGA*
A4	*Akkermansia*	9	47.53 - 49.57	2	rs81410866; rs81410881	8.78	5.25×10^{-6}	4.58×10^{-2}	*ssc-mir-125b-1*; *ssc-mir-100*
A5	*Akkermansia*	15	99.10 - 101.10	1	rs80982646	7.94	3.04×10^{-6}	4.58×10^{-2}	*SLC39A10*
B1	*CF231*	4	119.91 - 121.91	1	rs319005051	7.07	5.72×10^{-6}	6.48×10^{-2}	
B2	*CF231*	X	112.48 - 114.50	3	rs329229283	6.34	4.10×10^{-6}	6.48×10^{-2}	*FGF13*; *ATP11C*
C1	*Phascolarctobacterium*	9	65.33 - 67.33	1	rs81223434	7.73	1.25×10^{-6}	5.68×10^{-2}	*SLC45A3*; *RAB7B*; *RAB29*; *NUCKS1*; *IKBKE*; *MAPKAPK2*
D1	*Prevotella*	8	3.81 - 5.81	1	rs326174858	10.88	7.79×10^{-7}	3.53×10^{-2}	*CYTL1*; *WFS1*; *MAN2B2*
E1	*SMB53*	10	18.51 - 22.05	5	rs344136854	14.25	4.92×10^{-9}	1.68×10^{-4}	*CAPN8*; *CAPN2*; SUSD4; *DENND1B*; *PTPRC*; ssc-mir-181b-1; ssc-mir-181a-1
E2	*SMB53*	10	53.89 - 55.89	1	rs341165563	8.87	3.15×10^{-5}	7.14×10^{-2}	*MALRD1*
E3	*SMB53*	11	28.22 - 33.50	14	rs80835110	10.42	1.37×10^{-6}	1.47×10^{-2}	*PCDH17*
F1	*Streptococcus*	9	23.45 - 25.66	3	rs319168851	6.15	3.11×10^{-5}	9.41×10^{-2}	*FAT3*
F2	*Streptococcus*	13	2.97 - 5.15	4	rs81310237	11.88	1.20×10^{-5}	9.41×10^{-2}	*PLCL2; GALNT15*; *RFTN1*
F3	*Streptococcus*	14	133.80 - 135.80	1	rs337448241	10.51	3.46×10^{-6}	9.41×10^{-2}	*CTBP2*; *UROS*
F4	*Streptococcus*	15	25.15 - 27.87	9	rs331341379	8.72	1.12×10^{-5}	9.41×10^{-2}	*ERCC3*; *BIN1*; *MAP3K2*
F5	*Streptococcus*	18	44.25 - 46.25	1	rs334064749	7.7	2.26×10^{-5}	9.41×10^{-2}	*ssc-mir-196b-1*

[a]Chromosome; [b]Number of significant SNPs found in the region (FDR ≤ 0.1)

Akkermansia

The relative abundance of *Akkermansia*, a genus of the *Verrucomicrobia* phylum, was significantly associated with polymorphisms in five chromosomic regions: SSC3, SSC6, SSC7, SSC9, and SSC15 (Table 1). Within the SSC3 region (1.03-3.04 Mb), two candidate genes have been proposed, caspase recruitment domain family member 11 (*CARD11*) and carbohydrate sulfotransferase 12 (*CHST12*). CARD11 is necessary for T helper 17 cells differentiation which are involved in the adaptive immune system and protect the body against extracellular bacteria [39]. The *CHST12* gene is required for glycosaminoglycan biosynthesis [40]. Glycosaminoglycans are also called mucopolysaccharides and are often found in the mucin layer together with glycans and sialic acid [41]. The most common species of the *Verrucomicrobia* phylum found in the gut, *Akkermansia muciniphila*, colonizes the mucus layer and it is a known mucin degrader [42]. The regulation of host genes related to glycosaminoglycans biosynthesis probably has a direct effect in the occurrence of mucin degrading bacteria. Studying further this candidate gene may help to select a genetic variant that enriches the presence of *A. muciniphila*, since this species is beneficial to the host by restoring gut barrier function and helps reducing obesity [43]. In SSC6 (102.46 Mb), the only significant SNP (rs81390429, *p*-value=6.06×10^{-6}) explained a 25% of the variance in the abundance of the *Akkermansia* genus. The candidate gene found in this SSC6 region, *TGIF1* (TGFB induced factor homeobox 1), encodes for a protein that contributes to the adaptive immunity favouring the response of T follicular helper cells [44]. Additionally, two candidate genes were proposed for the *Akkermansia* spp. abundance in the SSC7 region (112.61-114.61 Mb): *CHGA* (chromogranin A) and *LGMN* (legumain). In humans, faecal levels of CHGA were associated with 61 different bacterial species including *A. muciniphila*, which was negatively associated with CHGA [45]. CHGA plays a role in the innate immunity with its antimicrobial activity against bacteria [46], whereas LGMN is a cysteine protease that also has antimicrobial activity, as well as it is involved in the antigen-presenting process and Toll-like receptors (TLRs) activation [47]. Thus, variations in the *CHGA* or *LGMN* genes should be affecting the microbiota composition based on the bacterial resistance to their antimicrobial activity. Inside the SSC9 region (47.53-49.57 Mb), there were two

candidate microRNAs genes: *ssc-mir-125b-1* and *ssc-mir-100*. Both microRNAs are involved with the adaptive immune system: *miR-125b-1* inhibits B cell differentiation [48], while *miR-100* inhibits T cell proliferation and differentiation [49]. Therefore, polymorphisms or differences in expression of both, the *miR-125b-1* and *miR-100* genes may be associated with the abundance of *Akkermansia* spp. Finally, a significant SNP (rs80982646, *p*-value=3.04×10^{-6}) located at 100.1 Mb in SSC15 was also associated with the relative abundance of *Akkermansia* spp. In this region, we identified the *SLC39A10* (solute carrier family 39 member 10) gene which appears to be a good candidate to modulate the presence of *Akkermansia* spp., since positively regulates B cell receptor signalling pathway [50]. Hence, in accordance with our results, germ-free mice colonized with *A. muciniphila* showed an overexpression in genes related with the antigen presentation pathway and B and T cell maturation, implying its possible role as host immune system modulator [51].

CF231

The relative abundance of the *CF231* genus (a member of the *Paraprevotellaceae* family) was associated to genetic variations in two regions along the pig genome in SSC4 and SSCX (Table 1). While no candidate genes were found in SSC4 at 120.91 Mb, the SSCX region (112.48-114.50 Mb) contained the ATPase phospholipid transporting 11C (*ATP11C*) and the fibroblast growth factor 13 (*FGF13*) genes. The ATP11C protein is involved in B cell differentiation past the pro-B cell stage, thus, defects in ATP11C led to a lower number of B cells and an impairment in their differentiation [52]. Changes in the *ATP11C* gene may cause species-specific tolerance through the adaptive immune system and transport. Additionally, ATP11C is also involved in the metabolism of cholestatic bile acids [53]. Intestinal content of cholesterol has the potential to shape the gut microbiome [54] and the *CF231* genus might be affected by the expression of these genes, since bile acids are catabolites of cholesterol. Interestingly, an enrichment of the *CF231* genus has been detected in experiments with high fat diet-induced hypercholesterolemic rats treated with cholesterol-lowering drugs [55]. On the other hand, the three significant SNPs of the SSCX region were located in an intron of the *FGF13* gene (Supplementary

Table S3). In this sense, FGF13 might be performing a similar function as FGF2 (fibroblast growth factor 2), repairing the intestinal epithelial damage [56].

Phascolarctobacterium

Six candidate genes found inside the SSC9 region (65.33-67.33 Mb) may be associated with the relative abundance of *Phascolarctobacterium* spp. (Table 1). *Phascolarctobacterium* is a Gram-negative genus commonly found in human faeces able to produce short chain fatty acids (SCFAs) [57]. SCFAs are absorbed and serve as a source of energy by colonocytes and peripheral tissue or can be use as substrates for lipogenesis, gluconeogenesis or regulation of cholesterol synthesis in the liver [58]. Interestingly, one of the candidate genes that could be modulating the abundance of *Phascolarctobacterium* spp. was the *SLC45A3* (solute carrier family 45 member 3). SLC45A3 is involved in the positive regulation of fatty acid biosynthetic process [59]. There were also two other candidate genes within SSC9 which encode GTPases that are members of the RAS oncogene family (*RAB7B* and *RAB29*). Under the induction of the lipopolysaccharides present in the Gram-negative cell wall, RAB7B promotes the degradation of toll like receptor 4 (TLR4) impairing the innate immune response by reducing the sensitivity of macrophages to lipopolysaccharides signalling [60]. Therefore, RAB7B may play an important role in the development of tolerance to Gram-negative commensal bacteria such as *Phascolarctobacterium* spp. The other GTPase, RAB29, is involved in bacterial toxin transport and is able to discriminate between *Salmonella enterica* serovars [61]. The positive regulation of insulin receptor signalling pathway by the *NUCKS1* (nuclear casein kinase and cyclin dependent kinase substrate 1) gene [62] located in this SSC9 region may also be modulating the abundance of the *Phascolarctobacterium* genus, since it has been described an enrichment of this genus in diabetic animal models treated with prebiotics to alleviate glucose intolerance [58]. Additionally, the two remaining candidate genes, *IKBKE* (inhibitor of nuclear factor kappa B kinase subunit epsilon) and *MAPKAPK2* (mitogen-activated protein kinase-activated protein kinase 2) might be associated with the microbiota composition because of their relationship with the immune system. IKBKE inhibits T cell responses [63]

and MAPKAPK2 regulates interleukin 10 [64] which is crucial to maintain the gut homeostasis [4].

Prevotella

Studies performed in humans have associated the presence of the *Prevotella* genus with a high intake of complex fibres in the diet [65]. In our animal material, *Prevotella* spp. represented 7.03% of the total composition at genus level (Supplementary Table S2). In this case, only the *rs326174858* SNP located at 4.81 Mb in SSC8 was significantly associated (p-value=7.79×10^{-7}) with the abundance of *Prevotella* spp. (Table 1). Three candidate genes were found in this SSC8 region, cytokine like 1 (*CYTL1*), wolframin ER transmembrane glycoprotein (*WFS1*), and mannosidase alpha class 2B member 2 (*MAN2B2*). *CYTL1* codes for a protein capable of chemoattracting macrophages and its activity is sensitive to *Bordetella pertussis* toxin [66]. A defect in the second gene, *WFS1*, produces insulin insufficiency, causing diabetes via pancreatic β cells failures [67]. Therefore, diabetic individuals would reduce glucose uptake in the gut epithelium [68]. In this sense, glucose might be more available for some bacteria species, producing changes in the overall microbiota composition. In accordance with this hypothesis, the abundance of *Prevotella* spp. was reduced in diabetic children when compared to healthy ones [69]. The encoded protein of the last candidate gene, *MAN2B2*, is implicated in the degradation of glycans [70]. Glycans are excreted into the intestine, including those in dietary plants, animal-derived, cartilage and tissue (glycosaminoglycans and N-linked glycans), and endogenous glycans from host mucus (O-linked glycans) [71]. The *Prevotella* genus contributes to the degradation of mucin and plant-based carbohydrates [72] and, therefore, it seems plausible that variations in a gene involved in the degradation of glycans could modulate the presence of *Prevotella* spp.

SMB53

The *SMB53* genus sequences found in swine compost were closely related with *Clostridium glycolicum* [73]. The abundance of the *SMB53* genus in our pig rectal samples accounted for 1.19% of the total number of annotated reads (Supplementary Table S2), and presented three significant associated regions,

two in SSC10 and one in SSC11 (Table 1). The first region of SSC10 (18.51-22.05 Mb) showed the most significant SNP (rs344136854, *p*-value=4.92×10^{-9}). Seven candidate genes have been identified inside this SSC10 region: calpain 2 (*CAPN2*) and 8 (*CAPN8*); sushi domain containing 4 (*SUSD4*); DENN domain containing 1B (*DENND1B*); protein tyrosine phosphatase, receptor type C (*PTPRC*); *ssc-mir-181a-1* and *ssc-mir-181b-1*. Calpains are a family of proteases that are able to perform various cellular functions depending on changes in intracellular Ca^{2+} levels [74]. For instance, an increase in Ca^{2+} in the intestinal porcine endothelial cells due to the *Clostridium perfringens* β-toxin triggers the calpain activation leading to intestinal cell death [75]. Therefore, polymorphisms in the *CAPN2* or *CAPN8* genes may confer resistance to *Clostridium* spp. avoiding endothelial cell death and so, increasing *Clostridia* abundance in the gut. Two of the significant SNPs of this SSC10 region were located in an intron of the *SUSD4* gene, whereas other significant SNP was also located in an intron of the *DENND1B* gene (Supplementary Table S3). *SUSD4*, *DENND1B* and *PTPRC* are genes related with the immune system: SUSD4 inhibits the complement system [76], DENND1B is a regulator of the T cell receptor signalling [77], and PTPRC is necessary for antigen receptor mediated signalling in lymphocytes [78]. The last two candidate genes in this first SSC10 region were both microRNAs from the *miR-181* family. Their depletion causes a lack of Natural Killer T cells in the thymus as well as defects in T and B cells development [79]. In the second SSC10 region (54.89 Mb), the only significant SNP (rs341165563, *p*-value=3.15×10^{-5}) was located in an intron of the *MALRD1* (MAM and LDL receptor class A domain containing 1) gene (Supplementary Table S3). This candidate gene is involved in bile acid synthesis regulation and is able to modify the gut microbiota [80]. Khan *et al*. [55] demonstrated an increase in the relative abundance of the *SMB53* genus in hypercholesterolemic rats treated with cholesterol-lowering drugs. Thus, further studies are needed to evaluate the modulation of the *SMB53* genus by the MALRD1 negative regulation of bile acid biosynthetic process. Additionally, the *SMB53* genus belongs to the *Clostridiaceae* family. Most members of this family have the capacity to consume gut mucus- and plant-derived saccharides like glucose [81]. Interestingly, recent studies performed by Horie *et al*. [82] have detected an enrichment of *SMB53* in

caecum of mice suffering type 2 diabetes, suggesting a possible role of this genus in the disease. The last region, in SSC11 (28.22-33.5 Mb), was comprised of 14 SNPs but, despite being the longest region observed (5.3 Mb), only one candidate gene (protocadherin 17, *PCDH17*) was proposed. Remarkably, the most significant SNP (rs80835110, *p*-value=1.37×10^{-6}) was located in an intron of *PCDH17* (Supplementary Table S3). PCDH17 may play a role in the colon similar to protocadherin 1 (PCDH1), acting as a physical barrier in the airway epithelial cells [83].

Streptococcus

There are five regions within the pig genome associated to the presence of *Streptococcus* spp., SSC9, SSC13, SSC14, SSC15, and SSC18 (Table 1). In the SSC9 region (23.45-25.66 Mb), the protein encoded by the FAT atypical cadherin 3 (*FAT3*) gene may be forming epithelial junctions that can be broken down by *Streptococcus* spp. [84]. In the SSC13 region (2.97-5.15 Mb), three candidate genes were found: phospholipase C like 2 (*PLCL2*), polypeptide N-acetylgalactosaminyltransferase 5 (*GALNT15*), and raftlin, lipid raft linker 1 (*RFTN1*). Four significant SNPs were located in intronic regions of the *PLCL2* gene (Supplementary Table S3). PLCL2 increases the thresholds of B cell activation [85] and so, it may modulate the tolerance of the adaptive immune system. On the other hand, GALNT15 belongs to a family of proteins that are able to produce O-linked glycosylation in the mucin [86] and hence, the variations on the *GALNT15* gene might affect some mucin dwellers like *Streptococcus* spp. [87]. Additionally, it is also interesting to highlight a possible link between the *RFTN1* gene, involved in the formation and/or maintenance of lipid rafts [88], and the abundance of the *Streptococcus* genus. The lipid rafts are microdomains located in the membrane surface of the cell that play an important role in cellular signaling and membrane trafficking of T and B lymphocytes [88, 89]. Furthermore, lipid rafts are also mediators of innate immune recognition of bacteria [90]. The possible association of RFTN1 with the abundance of *Streptococci* needs further attention, since some species of *Streptococcus* are known to hijack these lipid rafts to enter the host cell causing disease [91]. Two candidate genes were found in the SSC14 region (133.8-135.8 Mb): *CTBP2* (C-terminal binding protein 2) and

UROS (uroporphyrinogen III synthase). The *CTBP2* gene was associated in pigs with a susceptibility to develop a bacterial respiratory disease [92]. The *UROS* gene is involved in the metabolism of porphyrins including heme and uroporphyrinogen III biosynthetic processes [93]. Iron in mammals is incorporated into heme; an essential component of the hemoglobin, which can be acquired by bacterial pathogens as a nutritional iron source. Several *Streptococci* species that are pathogenic to humans and animals, namely *S. pyogenes*, *S. pneumoniae* and *S. suis*, contain cell wall heme-binding proteins that allow them to scavenge heme from host's hemoglobin as a source of iron acquisition [94, 95]. Additionally, the group B *Streptococci* are able to respire in the presence of heme, enhancing resistance to oxidative stress and improving their survival [96]. Our results suggest that the *UROS* gene may modulate the presence of *Streptococcus* spp. making these animals more susceptible to *Streptococci* colonization. A total of three candidate genes were identified in the SSC15 region (25.15-27.87 Mb): *ERCC3* (ERCC excision repair 3, TFIIH core complex helicase subunit), *BIN1* (bridging integrator 1), and *MAP3K2* (mitogen-activated protein kinase kinase kinase 2). The *ERCC3* gene expression was downregulated in human gastric cells after the infection with *Helicobacter pylori* [97]. In the same direction as the aforementioned *GALNT15* gene, the *BIN1* gene might modulate the abundance of mucin dweller bacteria like *Streptococcus* spp. because of the attenuation of *BIN1* favours the intestinal barrier function [98]. The protein encoded by the last candidate gene of this SSC15 region, *MAP3K2*, activates the toll like receptor 9 (TLR9) that recognizes CpG oligodeoxynucleotide motif in bacteria [99]. Finally, the last significant region (45.25 Mb) in SSC18 contained one microRNA, *ssc-mir-196b-1*, that was found upregulated in the duodenum of piglets that were resistant to *Escherichia coli* infection [100].

Conclusion

This report identifies associations between the pig genome and the relative abundance of six genera (*Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53* and *Streptococcus*). Most of the candidate genes found in the 17 associated regions of the pig genome encode for proteins that are involved in

the host defence system, including the immune system, physical barriers such as the mucin layer or cell junctions, whereas other proteins participate in the metabolism of mucopolysaccharides or bile acids. Our results confirm the importance of host genomics in the modulation of the microbiota composition. Further studies are warranted to determine which genetic combinations favour the enrichment of beneficial bacteria, providing the individual with the best intestinal health to avoid the entrance of potential pathogens.

Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness (MINECO AGL2014-56369-C2 and AGL2017-82641-R). D. Crespo-Piazuelo was funded by a "*Formació i Contractació de Personal Investigador Novell*" (FI-DGR) Ph.D grant from the *Generalitat de Catalunya* (ECO/1788/2014) and by the PiGutNet COST Action (www.pigutnet.eu) for a Short Term Scientific Mission at the GABI laboratory (INRA, France) under the supervision of J. Estellé. Contract of L. Migura-Garcia was supported by INIA and the European Social Fund. L. Criado-Mesas was funded with a FPI grant from the AGL2014-56369-C2 project. M. Revilla was also funded by a FI-DGR (ECO/1639/2013). M. Ballester was financially supported by a "*Ramón y Cajal*" contract (RYC-2013-12573) from the Spanish Ministry of Economy and Competitiveness. We acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the "*Severo Ochoa Programme for Centres of Excellence in R&D*" 2016-2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics and the CERCA Programme / *Generalitat de Catalunya*.

The authors would like to thank the Mazafra S. L. slaughterhouse for providing access to the data and material used in this study and, specially, to Francisco Minero for the skilful veterinary assistance. We also acknowledge the contribution of Rita Benítez in the collection of samples and microbial DNA extractions.

Authors' contributions

JMF and AIF conceived and designed the experiments; JMF was the principal investigator of the project; this work is part of the PhD thesis of DCP co-supervised by MB and JMF; JMGC provided animal samples; DCP, MR, MM,

JMGC and AIF collected samples; DCP and MB tested the DNA extraction protocol; DCP and MM performed the microbial DNA extraction; LCM performed the pig genomic DNA extraction; AC genotyped the samples; DCP and JE analysed the data; DCP, LMG, MB, and JMF wrote the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. *Science* 2012; **336**: 1262–7.
2. Caballero S, Pamer EG. Microbiota-mediated inflammation and antimicrobial defense in the intestine. *Annu Rev Immunol* 2015; **33**: 227–56.
3. Lederberg J, McCray A. 'Ome Sweet 'Omics—a genealogical treasury of words. *Scientist* 2001; **15**: 8.
4. Kamada N, Seo S-U, Chen GY, Núñez G. Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol* 2013; **13**: 321–335.
5. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman D a. The application of ecological theory toward an understanding of the human microbiome. *Science* 2012; **336**: 1255–62.
6. Johansson ME V, Ambort D, Pelaseyed T, Schütte A, Gustafsson JK, Ermund A, et al. Composition and functional role of the mucus layers in the intestine. *Cell Mol Life Sci* 2011; **68**: 3635–41.
7. Johansson ME V, Larsson JMH, Hansson GC. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* 2011; **108 Suppl**: 4659–65.
8. Hooper L V, Stappenbeck TS, Hong C V, Gordon JI. Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol* 2003; **4**: 269–73.
9. Chu H, Mazmanian SK. Innate immune recognition of the microbiota

promotes host-microbial symbiosis. *Nat Immunol* 2013; **14**: 668–675.

10. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012; **486**: 207–14.
11. Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, et al. A reference gene catalogue of the pig gut microbiome. *Nat Microbiol* 2016; **1**: 16161.
12. Wang M, Donovan SM. Human microbiota-associated swine: current progress and future opportunities. *ILAR J* 2015; **56**: 63–73.
13. Estellé J, Mach N, Ramayo-Caldas Y, Levenez F, Lemonnier G, Denis C, et al. The influence of host's genetics on the gut microbiota composition in pigs and its links with immunity traits. *10th World Congr. Genet. Appl. to Livest. Prod. Vancouver, BC, Canada*. 2014.
14. 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**: 1637–1644.
15. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, et al. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat Genet* 2016; **48**: 1413–1417.
16. Goodrich JK, Davenport ER, Clark AG, Ley RE. The Relationship Between the Human Genome and Microbiome Comes into View. *Annu Rev Genet* 2017; **51**: 413–433.
17. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, et al. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol* 2015; **16**: 191.
18. Goodrich JK, Davenport ER, Waters JL, Clark AG, Ley RE. Cross-species comparisons of host genetic associations with the microbiome. *Science* 2016; **352**: 532–5.
19. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* 2011; **9**: 279–90.
20. Estellé J, Mach N, Ramayo-Caldas Y, Levenez F, Lemonnier G, Denis C, et al. Host genetics influences gut microbiota composition in pigs. *36th Int. Soc. Anim. Genet. Conf. Dublin, Irel.* 2017.

21. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010; **7**: 335–6.
22. Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, et al. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ* 2014; **2**: e545.
23. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 2013; **10**: 57–9.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–10.
25. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol* 2006; **72**: 5069–5072.
26. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013; **8**: e61217.
27. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2016.
28. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2009. Springer-Verlag New York.
29. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr* 1957; **27**: 325–349.
30. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. 2nd ed. 1989. Cold Spring Harbor Laboratory Press, pp E3–E4.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.
32. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012; **44**: 821–4.
33. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical

and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 1995; **57**: 289–300.

34. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011; **2011**: bar030.
35. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010; **26**: 2069–2070.
36. Ramayo-Caldas Y, Mach N, Lepage P, Levenez F, Denis C, Lemonnier G, et al. Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J* 2016; **10**: 2973–2977.
37. Holman DB, Brunelle BW, Trachsel J, Allen HK. Meta-analysis To Define a Core Microbiota in the Swine Gut. *mSystems* 2017; **2**: e00004-17.
38. Zhao W, Wang Y, Liu S, Huang J, Zhai Z, He C, et al. The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments. *PLoS One* 2015; **10**: 1–13.
39. Molinero LL, Cubre A, Mora-Solano C, Wang Y, Alegre M-L. T cell receptor/CARMA1/NF-κB signaling controls T-helper (Th) 17 differentiation. *Proc Natl Acad Sci U S A* 2012; **109**: 18529–34.
40. Hiraoka N, Nakagawa H, Ong E, Akama TO, Fukuda MN, Fukuda M. Molecular cloning and expression of two distinct human chondroitin 4-O-sulfotransferases that belong to the HNK-1 sulfotransferase gene family. *J Biol Chem* 2000; **275**: 20188–96.
41. Ouwerkerk JP, de Vos WM, Belzer C. Glycobiome: bacteria and mucus at the epithelial interface. *Best Pract Res Clin Gastroenterol* 2013; **27**: 25–38.
42. Ottman N, Geerlings SY, Aalvink S, de Vos WM, Belzer C. Action and function of Akkermansia muciniphila in microbiome ecology, health and disease. *Best Pract Res Clin Gastroenterol* 2017; **31**: 637–642.
43. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, et al. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A* 2013; **110**: 9066–71.

44. Leber A, Abedi V, Hontecillas R, Viladomiu M, Hoops S, Ciupe S, et al. Bistability analyses of CD4+ T follicular helper and regulatory cells during Helicobacter pylori infection. *J Theor Biol* 2016; **398**: 74–84.
45. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 2016; **352**: 565–9.
46. Briolat J, Wu SD, Mahata SK, Gonthier B, Bagnard D, Chasserot-Golaz S, et al. New antimicrobial activity for the catecholamine release-inhibitory peptide from chromogranin A. *C Cell Mol Life Sci* 2005; **62**: 377–385.
47. Dall E, Brandstetter H. Structure and function of legumain in health and disease. *Biochimie* 2016; **122**: 126–50.
48. Gururajan M, Haga CL, Das S, Leu C-M, Hodson D, Josson S, et al. MicroRNA 125b inhibition of B cell differentiation in germinal centers. *Int Immunol* 2010; **22**: 583–92.
49. Negi V, Paul D, Das S, Bajpai P, Singh S, Mukhopadhyay A, et al. Altered expression and editing of miRNA-100 regulates iTreg differentiation. *Nucleic Acids Res* 2015; **43**: 8057–65.
50. Hojyo S, Miyai T, Fujishiro H, Kawamura M, Yasuda T, Hijikata A, et al. Zinc transporter SLC39A10/ZIP10 controls humoral immunity by modulating B-cell receptor signal strength. *Proc Natl Acad Sci U S A* 2014; **111**: 11786–91.
51. Derrien M, Van Baarlen P, Hooiveld G, Norin E, Müller M, de Vos WM. Modulation of Mucosal Immune Response, Tolerance, and Proliferation in Mice Colonized by the Mucin-Degrader Akkermansia muciniphila. *Front Microbiol* 2011; **2**: 166.
52. Yabas M, Teh CE, Frankenreiter S, Lal D, Roots CM, Whittle B, et al. ATP11C is critical for the internalization of phosphatidylserine and differentiation of B lymphocytes. *Nat Immunol* 2011; **12**: 441–9.
53. Siggs OM, Schnabl B, Webb B, Beutler B. X-linked cholestasis in mouse due to mutations of the P4-ATPase ATP11C. *Proc Natl Acad Sci U S A* 2011; **108**: 7890–5.
54. Islam KBMS, Fukiya S, Hagio M, Fujii N, Ishizuka S, Ooka T, et al. Bile acid

is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology* 2011; **141**: 1773–81.

55. Khan TJ, Ahmed YM, Zamzami MA, Mohamed SA, Khan I, Baothman OAS, et al. Effect of atorvastatin on the gut microbiota of high fat diet-induced hypercholesterolemic rats. *Sci Rep* 2018; **8**: 662.
56. Song X, Dai D, He X, Zhu S, Yao Y, Gao H, et al. Growth Factor FGF2 Cooperates with Interleukin-17 to Repair Intestinal Epithelial Damage. *Immunity* 2015; **43**: 488–501.
57. Wu F, Guo X, Zhang J, Zhang M, Ou Z, Peng Y. Phascolarctobacterium faecium abundant colonization in human gastrointestinal tract. *Exp Ther Med* 2017; **14**: 3122–3126.
58. Zhang Q, Yu H, Xiao X, Hu L, Xin F, Yu X. Inulin-type fructan improves diabetic phenotype and gut microbiota profiles in rats. *PeerJ* 2018; **6**: e4446.
59. Shin D, Howng SYB, Ptáček LJ, Fu Y-H. miR-32 and its target SLC45A3 regulate the lipid metabolism of oligodendrocytes and myelin. *Neuroscience* 2012; **213**: 29–37.
60. Wang Y, Chen T, Han C, He D, Liu H, An H, et al. Lysosome-associated small Rab GTPase Rab7b negatively regulates TLR4 signaling in macrophages by promoting lysosomal degradation of TLR4. *Blood* 2007; **110**: 962–71.
61. Spanò S, Liu X, Galán JE. Proteolytic targeting of Rab29 by an effector protein distinguishes the intracellular compartments of human-adapted and broad-host Salmonella. *Proc Natl Acad Sci U S A* 2011; **108**: 18418–23.
62. Qiu B, Shi X, Wong ET, Lim J, Bezzi M, Low D, et al. NUCKS is a positive transcriptional regulator of insulin signaling. *Cell Rep* 2014; **7**: 1876–86.
63. Zhang J, Feng H, Zhao J, Feldman ER, Chen S-Y, Yuan W, et al. IκB Kinase ε Is an NFATc1 Kinase that Inhibits T Cell Immune Response. *Cell Rep* 2016; **16**: 405–418.
64. Ehlting C, Trilling M, Tiedje C, Le-Trilling VTK, Albrecht U, Kluge S, et al. MAPKAP kinase 2 regulates IL-10 expression and prevents formation of intrahepatic myeloid cell aggregates during cytomegalovirus infections. *J Hepatol* 2016; **64**: 380–389.

65. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* 2010; **107**: 14691–6.
66. Wang X, Li T, Wang W, Yuan W, Liu H, Cheng Y, et al. Cytokine-like 1 Chemoattracts Monocytes/Macrophages via CCR2. *J Immunol* 2016; **196**: 4090–9.
67. Bonnycastle LL, Chines PS, Hara T, Huyghe JR, Swift AJ, Heikinheimo P, et al. Autosomal dominant diabetes arising from a Wolfram syndrome 1 mutation. *Diabetes* 2013; **62**: 3943–50.
68. Ussar S, Haering M-F, Fujisaka S, Lutter D, Lee KY, Li N, et al. Regulation of Glucose Uptake and Enteroendocrine Function by the Intestinal Epithelial Insulin Receptor. *Diabetes* 2017; **66**: 886–896.
69. Murri M, Leiva I, Gomez-Zumaquero JM, Tinahones FJ, Cardona F, Soriguer F, et al. Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. *BMC Med* 2013; **11**: 46.
70. Venkatesan M, Kuntz DA, Rose DR. Human lysosomal alpha-mannosidases exhibit different inhibition and metal binding properties. *Protein Sci* 2009; **18**: 2242–51.
71. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 2012; **10**: 323–35.
72. Pajarillo EAB, Chae JP, Kim HB, Kim IH, Kang D-K. Barcoded pyrosequencing-based metagenomic analysis of the faecal microbiome of three purebred pig lines after cohabitation. *Appl Microbiol Biotechnol* 2015; **99**: 5647–56.
73. Guo Y, Zhu N, Zhu S, Deng C. Molecular phylogenetic diversity of bacteria and its spatial distribution in composts. *J Appl Microbiol* 2007; **103**: 1344–54.
74. Kumar V, Ali A. Targeting calpains: A novel immunomodulatory approach for microbial infections. *Eur J Pharmacol* 2017; **814**: 28–44.
75. Autheman D, Wyder M, Popoff M, D'Herde K, Christen S, Posthaus H. Clostridium perfringens beta-toxin induces necrostatin-inhibitable, calpain-dependent necrosis in primary porcine endothelial cells. *PLoS One* 2013;

8: e64644.

76. Holmquist E, Okroj M, Nodin B, Jirström K, Blom AM. Sushi domain-containing protein 4 (SUSD4) inhibits complement by disrupting the formation of the classical C3 convertase. *FASEB J* 2013; **27**: 2355–66.
77. Yang C-W, Hojer CD, Zhou M, Wu X, Wuster A, Lee WP, et al. Regulation of T Cell Receptor Signaling by DENND1B in TH2 Cells and Allergic Disease. *Cell* 2016; **164**: 141–155.
78. Hermiston ML, Xu Z, Weiss A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol* 2003; **21**: 107–37.
79. Henao-Mejia J, Williams A, Goff LA, Staron M, Licona-Limón P, Kaech SM, et al. The microRNA miR-181 is a critical cellular metabolic rheostat essential for NKT cell ontogenesis and lymphocyte development and homeostasis. *Immunity* 2013; **38**: 984–97.
80. Li T, Chiang JYL. Bile acids as metabolic regulators. *Curr Opin Gastroenterol* 2015; **31**: 159–65.
81. Wüst PK, Horn MA, Drake HL. Clostridiaceae and Enterobacteriaceae as active fermenters in earthworm gut content. *ISME J* 2011; **5**: 92–106.
82. Horie M, Miura T, Hirakata S, Hosoyama A, Sugino S, Umeno A, et al. Comparative analysis of the intestinal flora in type 2 diabetes and nondiabetic mice. *Exp Anim* 2017; **66**: 405–416.
83. Kozu Y, Gon Y, Maruoka S, Kazumichi K, Sekiyama A, Kishi H, et al. Protocadherin-1 is a glucocorticoid-responsive critical regulator of airway epithelial barrier function. *BMC Pulm Med* 2015; **15**: 1–12.
84. Xu H, Sobue T, Bertolini M, Thompson A, Dongari-Bagtzoglou A. Streptococcus oralis and Candida albicans Synergistically Activate μ-Calpain to Degrade E-cadherin From Oral Epithelial Junctions. *J Infect Dis* 2016; **214**: 925–34.
85. Takenaka K, Fukami K, Otsuki M, Nakamura Y, Kataoka Y, Wada M, et al. Role of phospholipase C-L2, a novel phospholipase C-like protein that lacks lipase activity, in B-cell receptor signaling. *Mol Cell Biol* 2003; **23**: 7329–38.
86. Clausen H, Bennett EP. A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-

linked glycosylation. *Glycobiology* 1996; **6**: 635–46.

87. Homer KA, Patel R, Beighton D. Effects of N-acetylglucosamine on carbohydrate fermentation by Streptococcus mutans NCTC 10449 and Streptococcus sobrinus SL-1. *Infect Immun* 1993; **61**: 295–302.
88. Saeki K, Miura Y, Aki D, Kurosaki T, Yoshimura A. The B cell-specific major raft protein, Raftlin, is necessary for the integrity of lipid raft and BCR signal transduction. *EMBO J* 2003; **22**: 3015–26.
89. Alonso MA, Millán J. The role of lipid rafts in signalling and membrane trafficking in T lymphocytes. *J Cell Sci* 2001; **114**: 3957–65.
90. Triantafilou M, Miyake K, Golenbock DT, Triantafilou K. Mediators of innate immune recognition of bacteria concentrate in lipid rafts and facilitate lipopolysaccharide-induced cell activation. *J Cell Sci* 2002; **115**: 2603–11.
91. Toledo A, Benach JL. Hijacking and Use of Host Lipids by Intracellular Pathogens. *Microbiol Spectr* 2015; **3**: 637–666.
92. Huang X, Huang T, Deng W, Yan G, Qiu H, Huang Y, et al. Genome-wide association studies identify susceptibility loci affecting respiratory disease in Chinese Erhualian pigs under natural conditions. *Anim Genet* 2017; **48**: 30–37.
93. Tsai SF, Bishop DF, Desnick RJ. Human uroporphyrinogen III synthase: molecular cloning, nucleotide sequence, and expression of a full-length cDNA. *Proc Natl Acad Sci U S A* 1988; **85**: 7049–53.
94. Eichenbaum Z, Muller E, Morse SA, Scott JR. Acquisition of iron from host proteins by the group A streptococcus. *Infect Immun* 1996; **64**: 5428–9.
95. Wan Y, Zhang S, Li L, Chen H, Zhou R. Characterization of a novel streptococcal heme-binding protein SntA and its interaction with host antioxidant protein AOP2. *Microb Pathog* 2017; **111**: 145–155.
96. Franza T, Delavenne E, Derré-Bobillot A, Juillard V, Boulay M, Demey E, et al. A partial metabolic pathway enables group b streptococcus to overcome quinone deficiency in a host bacterial community. *Mol Microbiol* 2016; **102**: 81–91.
97. Chiou CC, Chan CC, Sheu DL, Chen KT, Li YS, Chan EC. Helicobacter pylori infection induced alteration of gene expression in human gastric cells. *Gut* 2001; **48**: 598–604.

98. Chang MY, Boulden J, Valenzano MC, Soler AP, Muller AJ, Mullin JM, et al. Bin1 attenuation suppresses experimental colitis by enforcing intestinal barrier function. *Dig Dis Sci* 2012; **57**: 1813–21.

99. Wen M, Ma X, Cheng H, Jiang W, Xu X, Zhang Y, et al. Stk38 protein kinase preferentially inhibits TLR9-activated inflammatory responses by promoting MEKK2 ubiquitination in macrophages. *Nat Commun* 2015; **6**: 7167.

100. Wu Z, Qin W, Wu S, Zhu G, Bao W, Wu S. Identification of microRNAs regulating Escherichia coli F18 infection in Meishan weaned piglets. *Biol Direct* 2016; **11**: 59.

Supplementary Information

Supplementary Information S1. Full description of the 16S rRNA gene amplification and sequencing.

Supplementary Table S1. Means for the relative abundance of the 18 phyla found in rectal contents of 285 pigs. %_Presence indicates the percentage of the pigs where these phyla were found.

Supplementary Table S2. Means for the relative abundance of the 101 genera found in rectal contents of 285 pigs. %_Presence indicates the percentage of the pigs where these genera were found.

Supplementary Table S3. Description of the 52 significant genera-associated SNPs and their predicted consequences with the Variant Effect Predictor tool [35] (Ensembl release 92). MAF indicates the minor allele frequency.

4. GENERAL DISCUSSION

Animal breeding appeared in the 1700's and since then, it has been used to systematically select animals to be the progenitors of the next generation in order to fix a set of predefined characteristics in a population (Oldenbroek and van der Waaij, 2014). However, these selection goals are usually complex traits that are determined by a large number of unlinked *loci* with small effects, following the infinitesimal genetic model of inheritance (Fisher, 1918).

In domestic animals, the genetics of complex traits has usually been studied without identifying the genes involved (Goddard and Hayes, 2009). Selective breeding is based on the genetic merit (i.e., estimated breeding values) of an animal as a parent. This genetic merit has been calculated using phenotypic records, heritability and pedigree, and more recently, genetic markers (Meuwissen and Goddard, 1996). In this regard, the appearance of high-throughput SNP genotyping arrays has facilitated the discovery of causal genes and mutations causing monogenic Mendelian traits and the identification of some genes for complex traits, which remains a challenging task (Andersson and Georges, 2004).

Meat quality is a complex trait subjected to consumer preferences and it is determined by a list of factors such as food safety, animal welfare, FA composition and sensory aspects (Wood *et al.*, 1999; Webb and O'Neill, 2008; Barendse, 2014). Meat quality may be determined by a large number of *loci* of small effect and few genes of moderate effect (Hayes and Goddard, 2001). In addition, selection based on estimated breeding values can be a slow process if the trait can only be measured after the death of the animal (e.g., meat quality) (Goddard and Hayes, 2009). Therefore, the identification of the genes associated with meat quality would be useful to select animals carrying the favourable alleles (Meuwissen and Goddard, 1996).

Apart from alleles and genetic markers, gene expression levels, epigenetic variations and microbiota compositions can also be incorporated to predictive models for animal breeding, which may improve the prediction accuracy of the overall model (Mackay *et al.*, 2009; Camarinha-Silva *et al.*, 2017; Varona, 2017).

In this PhD thesis, we described genomic regions and variants associated with FA composition in muscle and backfat tissues in pig, as well as providing a

description of the microbiota found in the pig gut and the interaction with its host. In the following sections we discuss our results highlighting the identification of candidate genes associated with FA composition and gut microbiota, and the possible role of gut microbiota in the FA composition determination.

4.1 Genetic variants and candidate genes associated with fatty acid metabolism in muscle and adipose tissue

QTL mapping aims to identify chromosomal regions that can be associated with the variation in phenotypic traits. The appearance of high-throughput genotyping arrays with thousands of SNPs that are evenly distributed along the pig genome made possible the development of GWAS analysis (Goddard and Hayes, 2009). In this context, GWAS analyses have helped to reduce QTL intervals and to find causal mutations. GWAS for FA composition have been performed in different populations of crossed and purebred pigs such as Duroc, Landrace, Large White, and Erhualian (Ramayo-Caldas *et al.*, 2012a; Corominas *et al.*, 2013b; Muñoz *et al.*, 2013; B. Yang *et al.*, 2013; Zhang *et al.*, 2016a; Ros-Freixedes *et al.*, 2016; Zhang *et al.*, 2016b; Sato *et al.*, 2017; van Son *et al.*, 2017; Viterbo *et al.*, 2018; Zappaterra *et al.*, 2018). In these studies, strong candidate genes have been proposed to explain the QTLs for the FA composition of adipose tissue and muscle. The *ELOVL3* (ELOVL fatty acid elongase 3)*, ELOVL6*, *ELOVL7*, *FADS2*, *FASN* and *SCD* genes were found in common among our study and these previous GWAS (Table 4.1).

Table 4.1. Strong candidate genes to explain the QTLs for the FA composition of adipose tissue and muscle that were obtained through GWAS and were found in common with the present study in the merged dataset.

Reference	Population	Tissue	Candidate genes
(Ramayo-Caldas *et al.*, 2012a)	BC1_LD	LD	*ELOVL7*
(B. Yang et al., 2013)	Sutai, F_2	AF, LD	*ELOVL7*, *SCD*
(Muñoz et al., 2013)	BC1_LD	BF, LD	*ELOVL6*
(Zhang et al., 2016a)	DLY, Erhualian, Laiwu	LD	*ELOVL6*, *ELOVL7*, *FASN*, *SCD*
(Ros-Freixedes et al., 2016)	Duroc	GM, LD, SF	*SCD*
(Zhang et al., 2016b)	DLY, Erhualian, F2, Laiwu, Sutai	LD	*ELOVL6*, *ELOVL7*, *FADS2*, *FASN*, *SCD*
(Sato et al., 2017)	Duroc	BF, LD	*FASN*, *SCD*
(van Son et al., 2017)	Duroc, Landrace	SF	*ELOVL6*, *SCD*
(Viterbo et al., 2018)	Duroc	LD	*ELOVL3*, SCD
(Zappaterra et al., 2018)	Large White	BF	*ELOVL3*, *ELOVL6*, SCD
Present study	3BCs (BC1_DU, BC1_LD, BC1_PI)	BF, LD	*ELOVL3, ELOVL6*, *ELOVL7*, *FADS2*, *FASN*, *SCD*

AF: Abdominal fat; BC1_DU: (Iberian × Duroc) × Duroc; BC1_LD: (Iberian × Landrace) × Landrace; BC1_PI: (Iberian × Pietrain) × Pietrain; BF: Backfat; DLY: Duroc × (Landrace × Yorksire); F_2: White Duroc × Erhualian F_2; GM: *Gluteus medius*; LD*: Longissimus dorsi*; SF: Subcutaneous fat

In 2003, the first QTL genome scan for the FA composition of backfat in pigs was performed by the IBMAP project (Clop *et al.*, 2003). QTLs were identified on SSC4, SSC6, SSC8, SSC10, and SSC12 in a F_2 Iberian × Landrace cross using microsatellites. Later on, SNPs of the PorcineSNP60K BeadChip from *Illumina* that covered the whole pig genome were used to perform GWAS on the *Sscrofa9* assembly for the FA composition of IMF in 144 BC1_LD pigs (Ramayo-Caldas *et al.*, 2012a). A total of 43 regions that contained 813 trait-associated SNPs were identified.

In our study, GWAS was performed using a total of 38,424 SNPs mapped on the *Sscrofa11.1* assembly and the 60 phenotypic traits related with backfat and IMF FA composition (percentages, indices, and ratios) in three different pig crosses (n=441). These three different backcrosses were based on the Iberian breed: (Iberian × Landrace) × Landrace (BC1_LD, n=160), (Iberian × Duroc) × Duroc (BC1_DU, n=143), and (Iberian × Pietrain) × Pietrain (BC1_PI, n=138). This study has incorporated new data, because the BC1_DU and BC1_PI pigs had not been previously used in other studies. In addition, the increment of the sample size in this study is advantageous for GWAS analysis because it allows to improve the identification of meaningful associations, especially if the genetic architecture of a trait is formed by common variants of small effect (Korte and Farlow, 2013). Furthermore, the use of different populations tend to increase the accuracy of the genomic regions found because of the reduction of the long-range linkage disequilibrium regions (Goddard and Hayes, 2009).

In backfat, 96 significant associated SNPs located in nine *Sus scrofa* chromosomal regions of SSC1, SSC2, SSC4, SSC6, SSC8, SSC10, SSC12, and SSC16 were significantly associated (FDR≤0.1) with 12 traits: myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1(n-7)), 7-hexadecenoic acid (C16:1(n-9)), oleic acid (C18:1(n-9)), linoleic acid (C18:2(n-6)), dihomo-γ-linolenic acid (C20:3(n-6)), MUFA, PUFA, MUFA/PUFA, PUFA/SFA, and the gondoic acid/arachidic acid (C20:1(n-9)/C20:0) ratio. In these chromosomal regions, 33 candidate genes were located and proposed to explain the variations in the 12 backfat traits. Of these 33 candidate genes, four lipid metabolism-related genes (*ELOVL6*, *ELOVL7*, *FADS2*, *FASN*) were the most relevant as they have

been commonly described in GWAS for FA composition in adipose tissue and muscle in pigs (Table 4.1).

In the 0-10.77 Mb region of SSC2, the *FADS2* gene was suggested as a functional candidate gene to explain the variation in seven traits (C16:1(n-9), C18:1(n-9), C18:2(n-6), MUFA, PUFA, MUFA/PUFA, and PUFA/SFA). FADS2 desaturates the essential FAs (C18:2(n-6) and C18:3(n-3)) and other FAs (Figure 4.1), being able to modulate MUFA and PUFA content in backfat (Guillou *et al.*, 2004; Rioux *et al.*, 2015). This SSC2 QTL was previously described in the BC1_LD pigs for the percentages of C16:1(n-9), C18:2(n-6), C18:3(n-3), and PUFAS in backfat (Ballester *et al.*, 2016). In line with the previous study, we found a similar QTL region on SSC2 associated with five traits (C16:1(n-9), C18:2(n-6); PUFA, MUFA/PUFA and PUFA/SFA) when BC1_LD-specific GWAS was performed. In addition, this QTL signal was also found in the BC1_PI-specific GWAS for six traits (C17:0; C18:1(n-9), C18:2(n-6), MUFA, PUFA and MUFA/PUFA). In accordance with other GWAS performed for the intramuscular FA composition in Erhualian pigs (Zhang *et al.*, 2016b), *FADS2* was suggested as a candidate gene to explain the differences in two ratios, (C20:3(n-6)/C18:2(n-6) and C20:4(n-6)/C20:3(n-6). Recently, a polymorphism in the promoter of the *FADS2* gene (rs321384923) has been associated with the differences in the expression levels of *FADS2* in muscle and the desaturation efficiency of C18:2(n-6) to arachidonic acid (C20:4(n-6)) in muscle and adipose tissue of Duroc pigs (Gol *et al.*, 2018). Further analysis will be required to assess if rs321384923 or other polymorphism in the *FADS2* gene are the causal mutation of the SSC2 QTL, or the variation in backfat FA composition is produced by other genes of the SSC2 region, such as the fatty acid desaturase 1 and 3 (*FADS1* and *FADS3*) genes.

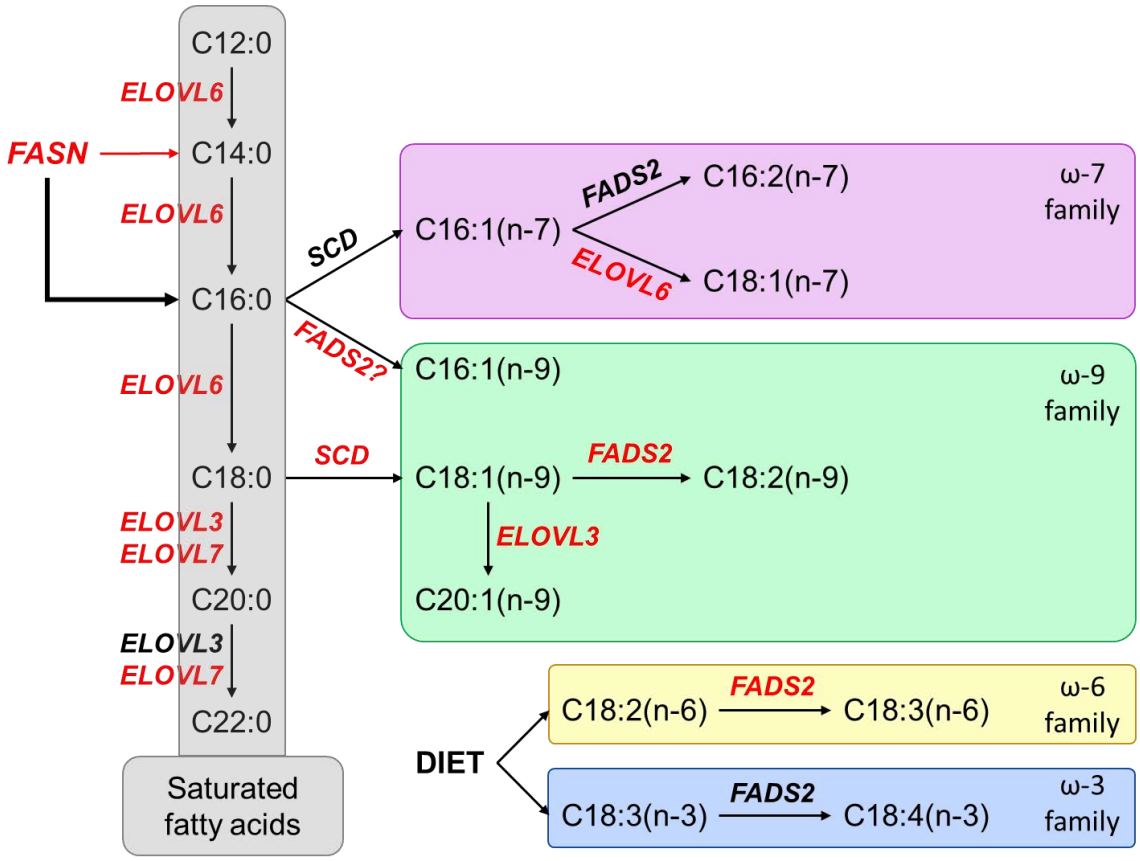


Figure 4.1. Summary of the fatty acid biosynthesis in pigs (adapted from Guillou *et al.*, 2010). Highlighted in red are the potential candidate genes to explain the variations in the fatty acid composition of adipose tissue and muscle. Arrows show the elongation and desaturation pathways of fatty acids whereby a potential candidate gene may be affecting a significant trait.

In the 108.40-116.41 Mb region of SSC8, the *ELOVL6* gene was proposed as a candidate gene to explain the variation of the percentages of C14:0, C16:0, and C16:1(n-7) in backfat. ELOVL6 can modulate the FA composition in muscle and in adipose tissue through the elongation of even C12-C16 SFAs and MUFAs (Matsuzaka and Shimano, 2009) (Figure 4.1). Our group previously reported the association between a polymorphism in the *ELOVL6* gene (*ELOVL6:c.-533C>T*) and the content of C16:0 and C16:1(n-7) in backfat and IMF of BC1_LD animals (Corominas *et al.*, 2013b). In a latter study, the *ELOVL6:c.-394G>A* polymorphism was proposed as the causal mutation to explain the QTL on SSC8 that modulates FA composition (Corominas *et al.*, 2015). This SSC8 QTL was later described in GWAS performed in different pig populations for the

percentages of C16:0, C16:1(n-7), C18:1(n-9) in adipose tissue and muscle suggesting the *ELOVL6* gene as the most promising candidate gene of this QTL (Zhang *et al.*, 2016b, 2016a; van Son *et al.*, 2017; Zappaterra *et al.*, 2018). In addition, the *ELOVL6* gene was differentially expressed in a RNA-Seq analysis performed in the adipose tissue of two phenotypically extreme groups of BC1_LD animals regarding their intramuscular FA composition (Corominas *et al.*, 2013a). To determine the association of the *ELOVL6* polymorphisms with an increasing number of animals, GWAS were reanalysed in the merged dataset including seven *ELOVL6*-SNPs already described in previous studies of our group (Corominas *et al.*, 2013b, 2015). Thus, *ELOVL6:c.-394G>A* was the most significant SNP of this QTL for the percentages of C14:0 (p-value=7.06×10^{-8}), C16:0 (p-value=2.43×10^{-12}) and C16:1(n-7) (p-value=2.6×10^{-7}) in backfat. Interestingly, when we performed the backcross-specific GWAS in the BC1_LD animals there was an overlap between the SSC8 QTLs for the FA composition (C16:0, C16:1(n-7), and C18:1(n-7)/C16:1(n-7)) in backfat (104.17-123.21 Mb) and in IMF (105.34-114.81 Mb), suggesting that *ELOVL6:c.-394G>A* affects the FA elongation in both tissues. However, the association with C14:0 was only found in backfat and it was not previously reported in the study performed in the BC1_LD pigs (Corominas *et al.*, 2015). Probably, the main reason of why the *ELOVL6:c.-394G>A* polymorphism was not previously found in association with the C14:0 content in backfat was the different sample size between studies. In the previous study, a reduced number of animals of BC1_LD were used (n=94) (Corominas *et al.*, 2015), while the present study was performed on 160 BC1_LD pigs. Furthermore, because the *ELOVL6:c.-394G>A* gene was not segregating in the BC1_DU and BC1_PI animals, no backcross-specific region on SSC8 was found in GWAS made for these backcrosses.

In the 0-1.91 Mb region of SSC12, the *FASN* was a promising candidate gene to explain the variations of the C14:0 content in backfat. FASN catalyses the synthesis of C16:0 and, to a lesser extent, C14:0 (Christie *et al.*, 1981) (Figure 4.1). However, C14:0 is mainly obtained from the diet (Legrand and Rioux, 2010) and can be synthesised through other different pathways (Rioux *et al.*, 2007). For example, other enzymes from the mitochondrial FAO such as LCAD or VLCAD may be producing the C16:0 shortening (Eaton *et al.*, 1996). Other GWAS

performed in Erhualian and Duroc pigs found a similar QTL region for the C14:0 abundance in adipose tissue and muscle (Zhang *et al.*, 2016a; Sato *et al.*, 2017). This QTL signal for the C14:0 abundance in backfat was also found in the BC1_DU-specific GWAS. Thus, the two most significant QTL signals found in the merged dataset for the C14:0 content in backfat were on SSC8 and SSC12, where the *ELOVL6* and *FASN* genes are located, respectively. Although the SSC8 QTL signal was mainly due to the BC1_LD animals and it was also associated with the percentages of C16:0 and C16:1(n-7). In addition, both candidate genes, *FASN* and *ELOVL6*, exhibited a higher expression in the adipose tissue of BC1_LD pigs with low PUFA content (Corominas *et al.*, 2013a). Therefore, a relationship can be established between *FASN* and *ELOVL6*, where FASN may be synthetizing *de novo* FAs that could be later elongated by ELOVL6.

In the 29.67-48.63 Mb region of SSC16, the *ELOVL7* gene was associated with the variation of the C20:1(n-9)/C20:0 ratio in backfat. The *ELOVL7* gene has been proposed as a candidate gene to explain a QTL for the variation in the abundance of C20:0 and the C20:1(n-9)/C20:0 ratio in the IMF of BC1_LD animals (Ramayo-Caldas *et al.*, 2012a). Moreover, GWAS performed in different populations also have found the association of *ELOVL7* with the percentages of C20:0 and the C20:1(n-9)/C20:0 ratio in adipose tissue and muscle (B. Yang *et al.*, 2013; Zhang *et al.*, 2016b, 2016a). ELOVL7 elongates C16-C20 FAs (Figure 4.1), especially FAs with a chain length of 18 carbon atoms (Naganuma *et al.*, 2011). With the aim of finding the causal mutation of the SSC16 QTL, three SNPs of the *ELOVL7* gene and two SNPs of the *PIK3R1* gene were included in the GWAS reanalysis for the C20:1(n-9)/C20:0 ratio in backfat in the merged dataset. Of these newly genotyped SNPs, the polymorphism located in the tenth exon of *ELOVL7* (*ELOVL7:c.*1432A>G*) was the most significant (p-value=1.8×10^{-7}). The *ELOVL7:c.*1432A>G* polymorphism also forms part of the 3'-UTR region of the *ELOVL7* gene, where miRNAs and transcription factors are able to bind and modulate gene expression (Barrett *et al.*, 2012), but no relevant binding motifs were found *in silico*. However, *ELOVL7:c.*1432A>G* was not the most significant SNP of the region, other two SNPs were, rs81297480 and rs81458871 (both p-values=3.2×10^{-8}). A similar QTL signal was found in the BC1_LD-specific GWAS for the C20:1(n-9)/C20:0 ratio in backfat, with a lower significance (p-

value=7.05×10^{-6}) than in the merged dataset GWAS (*p*-value=1.8×10^{-7}), probably because of the presence of some heterozygous BC1_PI individuals in the merged dataset. Therefore, *ELOVL7* was proposed as a potential candidate gene to explain the variations in the C20:1(n-9)/C20:0 ratio, although further studies are warranted to identify the causal mutation.

In IMF, 39 SNPs located in six regions of SSC4, SSC13, SSC14 and SSC17 were significantly associated (FDR≤0.1) with six traits: stearic acid (C18:0), arachidic acid (C20:0), eicosatrienoic acid (C20:3(n-3)), MUFA/SFA, the oleic acid/stearic acid (C18:1(n-9)/C18:0) ratio, and the arachidonic acid/dihomo-γ-linolenic acid (C20:4(n-6)/C20:3(n-6)) ratio. In addition, 17 candidate genes were proposed to explain the variations in the six IMF traits.

Of these 17 candidate genes, the *SCD* gene has been one of the most-referenced genes in pig lipid metabolism. In the 109.95-114.62 Mb region of SSC14, the *SCD* gene was associated with the C18:1(n-9)/C18:0 and MUFA/SFA ratios in IMF. In addition, this SSC14 region was associated with the same traits (C18:1(n-9)/C18:0 and MUFA/SFA) in the BC1_LD-specific GWAS. Polymorphisms in SCD have a strong effect on FA composition in swine IMF, especially in MUFA content (Ros-Freixedes *et al.*, 2016), and melting point of fat (Uemoto *et al.*, 2012). Near the *SCD* gene is located another candidate gene of this SSC14 QTL, the *ELOVL3* gene. Both genes participate in lipid metabolism, while SCD desaturates C18:0 into C18:1(n-9) (Figure 4.1), ELOVL3 elongates these C18 FAs and others participating in the synthesis of C20-C24 SFAs and MUFAs (Zadravec *et al.*, 2010). Moreover, the *SCD* gene was differentially expressed in the adipose tissue and muscle of BC1_LD animals with extreme differences in intramuscular FA composition (Corominas *et al.*, 2013a; Puig-Oliveras *et al.*, 2014). This pig SSC14 region has been associated through GWAS analysis with several FAs traits (C16:0, C16:1(n-7), C18:0, C18:1(n-9), SFAs and MUFAs) in adipose tissue and IMF (B. Yang *et al.*, 2013; Ros-Freixedes *et al.*, 2016; Zhang *et al.*, 2016b, 2016a; Sato *et al.*, 2017; van Son *et al.*, 2017; Viterbo *et al.*, 2018; Zappaterra *et al.*, 2018). Therefore, further studies are required to identify polymorphisms in *SCD* or *ELOVL3* that could be responsible for variations in the FA composition.

Using a larger number of animals, in the merged dataset GWAS in comparison to the backcross-specific GWAS, reduced the number of QTL regions found. In backfat, backcross-specific GWAS for the FA composition found 19 associated regions in BC1_LD, seven in BC1_PI, and five in BC1_DU, whereas more significant associated regions were found in the backcross-specific GWAS for the FA composition in IMF, 36 in BC1_PI, 33 in BC1_LD, and 23 in BC1_DU. Thus, one of the advantages of using three different populations were the increase in accuracy due to the reduction of the long-range linkage disequilibrium regions. However, as it has been pointed in the bibliography (Goddard and Hayes, 2009), the main drawback of using non-related populations was the loss of segregation for certain alleles disregarding the increase in the number of animals.

Aside from SNPs, other polymorphisms such as indels can be used as molecular markers. Indels tend to produce more severe consequences than SNPs, altering the reading frame of genes or changing the total number of amino acids of the proteins. In addition, indels can also modify gene expression, affecting backfat thickness (Ren *et al.*, 2012) and fat deposition (Zang *et al.*, 2016). In this context, we decided to identify indels from WGS data of two Iberian boars and five Landrace sows that were founders of the BC1_LD and we used a selection of the indels for genotyping and reanalysing the GWAS for the FA composition in IMF.

A total of 1,928,746 indels were found in common using three different programs for indel detection: *Dindel* (Albers *et al.*, 2011), *Genome Analysis Toolkit* (*GATK*) (McKenna *et al.*, 2010) and *SAMtools mpileup* (Li *et al.*, 2009). In accordance with the literature in pigs (Chen *et al.*, 2014; Molnár *et al.*, 2014; Kang *et al.*, 2015), deletions were more frequent than insertions, except for the most frequent indel found, the 1 bp long. In addition, sexual chromosomes (SSCX and SSCY) had lower density of indels than autosomes, probably due to the appearance of hemizygous recessive lethal mutations in males and the low recombination rate, only possible in females. The severity of the indels was predicted through the *VEP* tool, identifying 1,289 indels that may have a high impact on protein sequence and function.

Nine indels were selected to include them in GWAS based on the frequencies in the IBMAP founders (i.e., if the indel was present at extreme frequencies) and

based on the severity prediction. GWAS for the FA composition in IMF were performed in the merged dataset and in each one of the three backcrosses individually. No significantly associated regions were found for GWAS performed in the merged dataset, but one suggestive association was found for the BC1_PI-specific GWAS. The indel of the *C1QTNF12* gene (*C1QTNF12:c.557_559delCCG*) located on SSC6 was suggestively associated (*p*-value=1.77x10^{-5}, FDR=5.34x10^{-2}) with the percentage of eicosadienoic acid (C20:2(n-6)) in IMF. The protein encoded by the *C1QTNF12* gene functions as an adipokine that participates in glucose metabolism and obesity in mice (Enomoto *et al.*, 2011; Wei *et al.*, 2012). In addition, despite *C1QTNF12:c.557_559delCCG* was found at extreme frequencies in the founders, being the two Iberian boars homozygous for the deletion, no homozygous animals were found in the three backcrosses. Hence, *C1QTNF12:c.557_559delCCG* may have a higher effect on homozygous individuals and further studies in other populations will be required.

Finally, the indel (rs792025734) of the *GZMA* gene and the indel (rs709630954) of the *SAMD4B* gene were inside the 29.67-48.63 Mb region on SSC16 (for the C20:1(n-9)/C20:0 ratio in backfat. However, none of the two indels were associated with this ratio or any other FA, being *ELOVL7* the most promising candidate gene of this QTL region.

4.2 Characterization of the pig gut microbiota along the digestive tract and associations between host genome and microbiota composition

The microbiome interacts with the environment and can affect the host phenotype. The prediction accuracies of the statistical models for complex traits are improved if the microbiota composition is included (Camarinha-Silva *et al.*, 2017; Maltecca *et al.*, 2018). In this sense, another variable should be included in the phenotypic variance calculi (Phenotype = Genotype + Microbiome + Environment). In addition, host genetics also interacts with the microbiota composition, increasing the complexity of the model (Phenotype = Genotype +

Microbiome + (Genotype × Microbiome) + Environment) (Lu *et al.*, 2018). Therefore, the characterization of the microbiota composition and its interaction with the host genome need to be incorporated into the phenotypic analysis.

The main factors that can influence the microbiota composition can be divided into environmental and host factors (Figure 4.2). Pig gut microbiota can be modified due to different environmental factors, such as feed (Burrough *et al.*, 2015), housing (Dou *et al.*, 2017), and medication (Looft *et al.*, 2012). Conversely, other microbiota modifying factors are inherent of the host, some of which are age (Mach *et al.*, 2015; Zhao *et al.*, 2015), sex (Xiao *et al.*, 2016), gut region (Zhao *et al.*, 2015; Yang *et al.*, 2016; Kelly *et al.*, 2017), and breed or host genetics (Benson *et al.*, 2010; Camarinha-Silva *et al.*, 2017). Furthermore, the mother also influence the establishment of the microbiota (Mach *et al.*, 2015). The neonate piglet is exposed to a large number of microorganisms after the passage through the birth canal and later, through the contact with the skin and faeces of the mother, among others, which produces an early colonization of certain microorganisms (Katouli *et al.*, 1997; Thompson *et al.*, 2008). In the middle of environmental and host factors there is the health status of the host, which can affect drastically the microbiota composition if the host suffers a disease (Dou *et al.*, 2017; Kim and Isaacson, 2017). Nevertheless, technical factors such as DNA extraction methods can also be a source of variation in the observed microbiota composition (Wagner Mackenzie *et al.*, 2015).

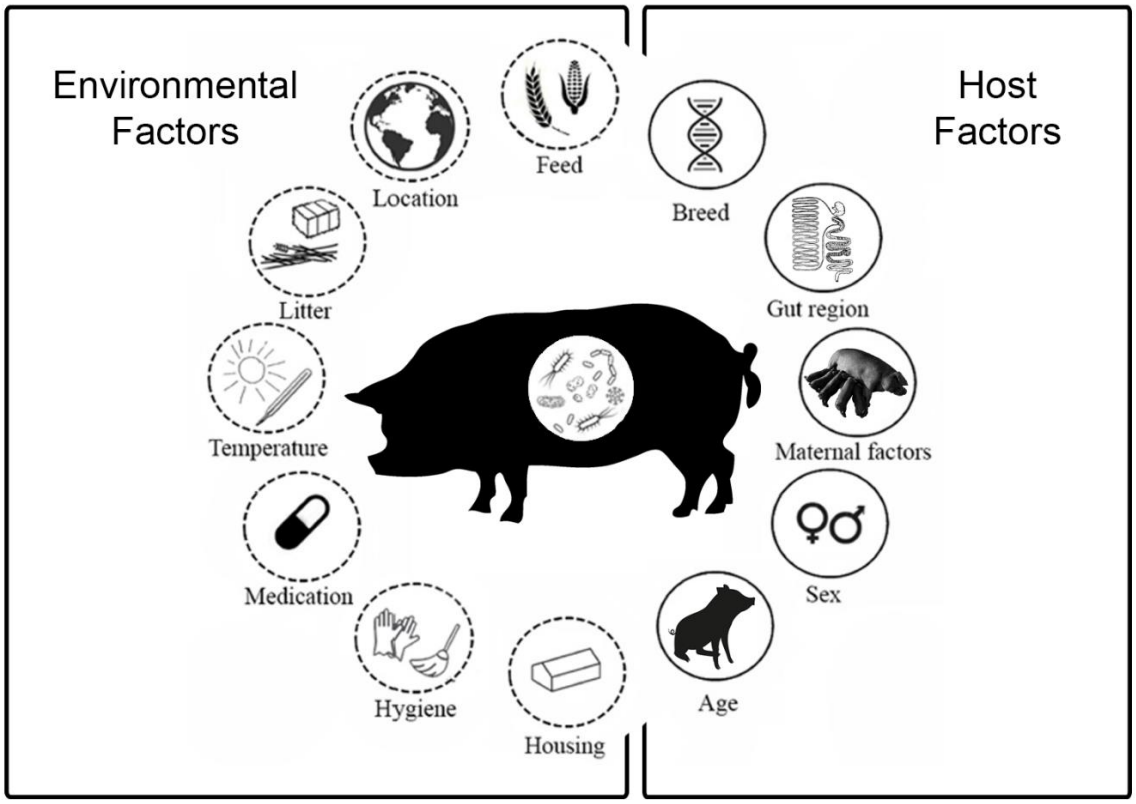


Figure 4.2. Factors that can affect the microbiota composition of the pig gut (adapted from Kers *et al.*, 2018).

In our first microbiota study, we have described the microbiota composition along the digestive tract of Iberian pigs to assess which gut section was more adequate to include in further phenotypic analysis. The availability of simple nutrients is lower in the large intestine than in the small intestine, while the number of microorganisms is higher (Donaldson *et al.*, 2016; Pereira and Berry, 2017). In this sense, colon samples were proved to be more informative due to the higher number of microorganisms present and the production of SCFAs (Liu *et al.*, 2012), which could affect FA composition in the host. Therefore, we selected the rectum in the second study of the microbiota, because it was the large intestine section easier to obtain in the abattoir for a high number of pigs.

To describe the differences among gut sections, we collected the luminal content of five gut sections (duodenum, jejunum, ileum, and proximal and distal colon) in thirteen castrated male Iberian purebred pigs at an age of 120 days and an average weight of 48.7 kg. Later on, the V3-V4 region of the 16S rRNA gene was

amplified and sequenced from these luminal contents and a total of 1,669 operational taxonomic units (OTUs) grouped in 179 genera were found. The most abundant genera in the duodenum and the jejunum were *Lactobacillus* and *Clostridium*, while *Streptococcus* spp. were more abundant in the ileum and *Prevotella* spp. were the most dominant in the colon. Hence, Bacteroidetes, Firmicutes and Proteobacteria were the most abundant phyla found along the digestive tract, which was previously reported by other studies of the luminal (Zhao *et al.*, 2015; Yang *et al.*, 2016) and mucosal (Kelly *et al.*, 2017) microbiota in pigs.

The higher α-diversity in the colon compared with the small intestine samples was in accordance with previous studies performed in Laiwu (Yang *et al.*, 2016) and Gloucestershire Old Spot (Kelly *et al.*, 2017) pigs. Conversely, the β-diversity decreased along the gut sections from the duodenum to the distal colon. Thus, the luminal contents of the colon samples were more similar and richer in species than the luminal contents of the small intestine samples. In this manner, the microbiota composition of the small intestine is possibly less stable than the large intestine microbiota due to the lower number of microorganisms found in the small intestine, the shorter transit time and the importance of adherence to mucus or tissue (Donaldson *et al.*, 2016; Pereira and Berry, 2017). In addition, several bacteria were only present in the colon and did not appear in the upper intestine sections. Apart from the rapid transit time in the small intestine (Donaldson *et al.*, 2016), other possible explanation of the differences in the microbiota composition among gut sections were the different environmental conditions, such as oxygen concentration (Marteyn *et al.*, 2011; Albenberg *et al.*, 2014), which make the settlement and growth of certain microorganisms less likely. Along with the differences in the oxygen concentration, which decreases in the colon, the availability of simple nutrients is also much lower in the large intestine than in the small intestine (Donaldson *et al.*, 2016; Pereira and Berry, 2017). Networks and functional prediction analyses based on genus abundances among gut sections suggested that, in the small intestine, there was a rapid internalization and conversion of the available simple carbohydrates for microbial proliferation and maintenance (Zoetendal *et al.*, 2012), whereas in the large intestine, anaerobic bacteria may compete for dietary fibre, such as cellulose or xylan (Bryant, 1959),

to use them as energy source, producing SCFAs (Liu *et al.*, 2012) that can be absorbed by the host (Brüssow and Parkinson, 2014).

The animals of the second study of the microbiota were 285 healthy male and female pigs from a commercial F1 (Duroc × Iberian) crossbred slaughtered at an average weight of 138.8 kg. The rectal contents of these pigs were collected and gut microbiota compositions were obtained through sequencing the V3-V4 region of the 16S rRNA gene. A total of 1,257 OTUs were found and distributed in 18 phyla and 101 genera. At phylum level, *Firmicutes* (45.36%) and *Bacteroidetes* (37.47%) were the two most abundant phyla, while *Prevotella* (7.03%) and *Treponema* (6.29%) were the two most abundant genera. In addition, the α-diversity mean was 5.58, while the reduced β-diversity indicated that the microbiota composition of the pigs was similar. When comparing studies of the gut microbiota with different environmental and host factors, results must be taken carefully. In our first study of the microbiota, samples were collected from the luminal content of five gut regions (duodenum, jejunum, ileum, and proximal and distal colon) in thirteen castrated male Iberian purebred pigs at an average weight of 48.7 kg, whereas in our second study of the microbiota, samples were collected from the luminal content of the rectum in 285 male and female pigs of a commercial Duroc × Iberian cross slaughtered at a higher average weight (138.8 kg). Despite these differences, due to the particular environmental conditions found in the large intestine, the microbiota composition found in the rectal content of the commercial crossbreed pigs was similar with the microbiota composition found in the proximal and distal colon of the Iberian purebreds (**Figure 4.3**). Furthermore, the α-diversities of rectum and proximal and distal colon calculated through the Shannon index were all above 5, representing that these regions were richer than the small intestine and in accordance with previous studies (Holman *et al.*, 2017).

As it was previously stated in the study of the gut microbiota along the digestive tract of the Iberian pigs, the different microbiota compositions found in the small and large intestines were probably due to the differences in the resources found along the digestive tract (Donaldson *et al.*, 2016; Pereira and Berry, 2017). In accordance with our two studies and several published articles (Zhao *et al.*, 2015; Yang *et al.*, 2016; Holman *et al.*, 2017), the abundance of Bacteroidetes in the

pig gut is usually higher in the large intestine (rectum and proximal and distal colon) than in the small intestine, whereas Firmicutes are predominant in the small intestine and, to a lesser extent, in the large intestine. Finally, the phylum Proteobacteria is the third most abundant, mainly in the small intestine.

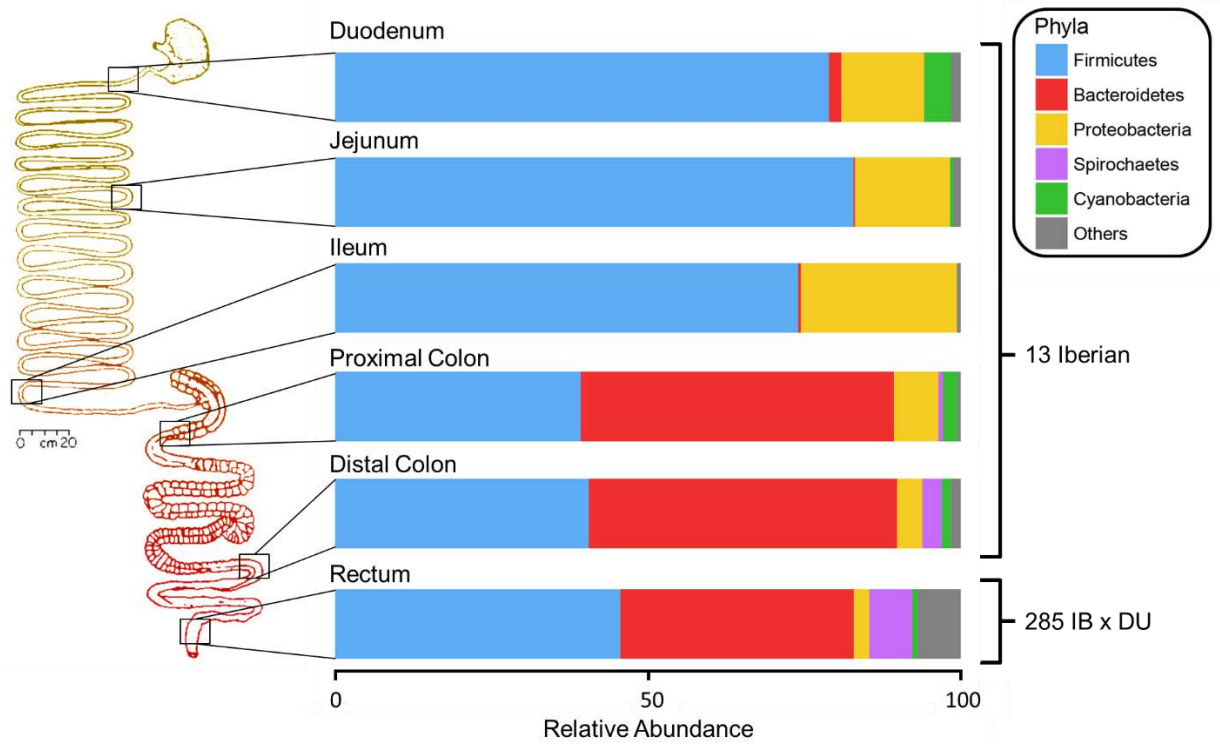


Figure 4.3. Stacked bar plot of the mean relative abundances (%) for the five most abundant phyla along the pig digestive tract. The microbiota composition of the first five regions (duodenum, jejunum, ileum, and proximal and distal colon) belong to the 13 Iberian population, whereas the microbiota composition of the last part, the rectum, belongs to the commercial population that is composed of 285 Iberian x Duroc pigs. Phyla are indicated by the colours in the upper right legend. The scheme of the pig digestive tract has been adapted from Varel, Yen and Kreikemeier, 1995.

The higher abundance of Bacteroidetes than Firmicutes found in the proximal and distal colon of the Iberian purebreds was reversed in the rectum of the commercial crossbreed pigs, resulting in a higher abundance of Firmicutes than Bacteroidetes. The different Bacteroidetes/Firmicutes ratio found between our two studies may be due to the different environmental factors whereby the studies were conducted (location and housing conditions) and the diverse host factors of the animals (sex, age, weight, fatness and breed). In addition, the feed was also

different between the two studies. The thirteen Iberian pigs were fed with a standard fodder based on maize and wheat, while the 285 crossbreed pigs were fed with a barley- and wheat-based commercial diet. The fodder of the thirteen Iberian pigs contained less barley than the commercial diet of the crossbreed pigs. Some species of the *Bacteroides*, *Prevotella*, *Clostridium*, and *Ruminococcus* genera are able to ferment barley derived β-glucans into SCFAs (Hughes *et al.*, 2008). The *Clostridium* and *Ruminococcus* genera belong to the Clostridiales order (Firmicutes phylum) and represented 39.66% of the total composition in the rectum, while Clostridiales were less abundant in the distal colon (33.64%). In addition, *Prevotella* spp. and *Bacteroides* spp. tend to be inversely related, and it has been shown that human gut samples with *Bacteroides* dominance are usually more associated with a Western diet, richer in amino acids and fats (Gorvitovskaia *et al.*, 2016). This is in accordance with our two studies, where the abundances of the *Prevotella* and *Bacteroides* genera in the distal colon were 35% and 0%, respectively, whereas in the rectum these abundances were 7.03% and 1.21%, respectively. However, species or strains of the *Prevotella* and *Bacteroides* genera may perform different functions and further studies, including whole-metagenome sequencing and metatranscriptomic analysis, will be required to determine the species or strain taxa and the main genes that were expressed.

One interesting aspect of our study of the microbiota was that the thirteen Iberian pigs were exposed to analogous environmental conditions, being raised and fed together. Therefore, if the environmental factors were minimized, the differences in the microbiota found among samples may have also been caused by host genetics, such as immune response-related genes (Fagarasan *et al.*, 2002).

There is controversy about the impact of host genetics in shaping gut microbiota. Host genetics has been proposed as an important factor in shaping gut microbiota composition (Goodrich *et al.*, 2014; Turpin *et al.*, 2016). The heritability of genus abundances in the pig gut range from low to high values (Estellé *et al.*, 2014; Camarinha-Silva *et al.*, 2017), unveiling a possible host genetic component in shaping the gut microbiota composition. Nonetheless, some studies in humans reported that environmental factors are predominant over host genetics (Spor *et al.*, 2011; Rothschild *et al.*, 2018), but conditions are difficult to standardize in

these studies. Conversely, production animals are better models to study or exclude environmental and host factors that shape gut microbiota, due to their similar diet, location, housing conditions and other factors, which can be standardized. In this regard, pigs are useful models to study human digestive diseases (Gonzalez *et al.*, 2015).

In this context, we performed a second study to analyse the role of host genetics in shaping the gut microbiota. Apart from the microbiota composition obtained from the rectal contents of the 285 commercial crossbred pigs, DNA was also obtained from the *Longissimus dorsi* muscle of these pigs to conduct the genotyping of 45,508 SNPs that covered the entire pig genome with a high-throughput method. Then, GWAS were made among pig genotypes and their gut microbiota composition. A total of 52 SNPs distributed in 17 regions along the pig genome (SSC3, SSC4, SSC6, SSC7, SSC8, SSC9, SSC10, SSC11, SSC13, SSC14, SSC15, SSC18 and SSCX) were associated with the relative abundance of six genera: *Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53*, and *Streptococcus*. A total of 38 candidate genes were proposed to explain the differences in the abundance of these six genera. Most of these candidate genes encoding proteins that play a role in the host defence, including the immune system and physical barriers (mucin layer or cell junctions), while other candidate genes encoding proteins that are more involved in the metabolism of mucopolysaccharides or bile acids. From the 38 suggested genes that may be modulating the pig microbiota composition, the most promising candidate genes were *UROS* (uroporphyrinogen III synthase) for the *Streptococcus* spp. abundance and *CHGA* (chromogranin A) for the *Akkermansia* spp. abundance.

The *UROS* gene participates in porphyrin metabolism, including heme biosynthesis (Tsai *et al.*, 1988). Heme complexes contain iron and are components of haemoglobin. This iron can be scavenged by several *Streptococci* species, including some pathogenic species such as *S. pyogenes*, *S. pneumoniae* and *S. suis* (Eichenbaum *et al.*, 1996; Wan *et al.*, 2017). Thus, *UROS* may be modifying the abundance of *Streptococcus* spp. through porphyrin metabolism, preventing or favouring *Streptococci* colonization and disease.

CHGA may be modulating the abundance of *Akkermansia* spp. through its antimicrobial activity and its role in innate immunity (Briolat *et al.*, 2005). Accordingly, CHGA levels in human faecal samples have been negatively associated with the most common species of the *Akkermansia* genus, *A. muciniphila* (Zhernakova *et al.*, 2016). This species, *A. muciniphila*, is a mucus layer colonizer and mucin degrader bacterium (Ottman *et al.*, 2017), which is able to restore gut barrier function and reduce obesity (Everard *et al.*, 2013).

In conclusion, in this study we identified genomic regions that may be influencing the gut microbiota composition through host-microbiota interactions in pigs, confirming the importance of host genomics in the modulation of the microbiota composition.

4.3 Future perspectives and challenges

GWAS performed on the merged dataset for the FA composition of backfat and IMF identified a total of 50 candidate genes. However, GWAS were only reanalysed with the genotyped polymorphisms found in the SSC8 region of the *ELOVL6* gene and in the SSC16 region of the *ELOVL7* gene. In addition, from all the genotyped SNPs, the *ELOVL6:c.-394G>A* polymorphism was the only one proposed as the causal mutation of the SSC8 QTL for the percentages of C14:0, C16:0, and C16:1(n-7) in backfat. Therefore, further validation analyses are required for identifying the causal mutations of the QTLs described in this study and genotyping them in the three backcrosses to perform association analysis. Furthermore, apart from the ten indels that were genotyped in the three backcrosses, other indels were found at extreme frequencies in the BC1_LD founders and predicted to have a high impact on protein sequence and function. Hence, further studies will be required to study the role of these indels in FA metabolism, especially for indels located inside GWAS regions in the merged dataset.

The studies included in this thesis have described the associations of pig genomics with meat quality traits and with microbiota genus composition in the gut. However, the association of gut microbiota with FA composition in muscle

and backfat has not been yet described in pigs. In this respect, apart from the microbiota composition of the 285 commercial Iberian × Duroc crossbreed pigs, our group has measured the FA composition in muscle and backfat for these animals as well as measures of the SCFA composition in their rectal content. Thus, future studies will analyse how the microbiota composition may affect muscle and backfat FA profile and the relationship between certain genera and the SCFA content in the gut.

The gut microbiota has been proposed as an important factor in the modulation of fatness in pigs (He *et al.*, 2016). Experiments transplanting the microbiota of obese humans to germ-free mice increased total body fat of mice compared with transplanting a human "lean" microbiota (Turnbaugh *et al.*, 2006). This experiment was later replicated in pigs producing the same results (Yan *et al.*, 2016). In addition, the mRNA expression profiles of lipid metabolism-related genes, such as *ACACA*, *FASN*, and *LPL* (lipoprotein lipase), in the skeletal muscle of obese pigs were replicated after transplanting the microbiota from obese pigs to germ-free mice, which also became obese, implicating that obese pig-derived microbiota enhances lipogenesis (Yan *et al.*, 2016). Therefore, microbiota composition is a transmissible trait that alters the capacity to harvest energy from the diet.

In this context, the Bacteroidetes/Firmicutes ratio in the gut has been traditionally associated with obesity and body fat mass in humans and pigs. Obese individuals tend to have more Firmicutes and less Bacteroidetes than lean individuals (Ley *et al.*, 2006; Million *et al.*, 2012; Pedersen *et al.*, 2013; Bashan *et al.*, 2016). A similar Bacteroidetes/Firmicutes ratio was obtained in one study using pig clones to reduce host genetic factors, where obesity was diet-induced (Pedersen *et al.*, 2013). Nevertheless, the Bacteroidetes/Firmicutes ratio is not an absolute indicator for obesity, other studies have reported increased levels of Bacteroidetes than Firmicutes in obese human individuals (Zhang *et al.*, 2009; Schwiertz *et al.*, 2010).

Dietary fibre is fermented by the gut microbiota into SCFAs (J. Yang *et al.*, 2013) (Figure 4.4), although some SCFAs can be originated directly from the diet (Høverstad and Midtvedt, 1986). The higher levels of Firmicutes in obese mice

generate an increment of the available energy that is obtained from the diet, mainly because Firmicutes produce more SCFAs than Bacteroidetes (Turnbaugh *et al.*, 2006; Murphy *et al.*, 2010; Scheithauer *et al.*, 2016). SCFAs are absorbed by the intestinal epithelium and nourish colonocytes (Cook and Sellin, 1998). Thereafter, SCFAs are transported to the liver (Bloemen *et al.*, 2009). These SCFAs, increase FAO in liver and muscle, while in adipose tissue decrease insulin sensitivity and fat storage (den Besten *et al.*, 2013; Scheithauer *et al.*, 2016). In addition, the microbiota promotes monosaccharide uptake into the intestinal epithelium, which lowers the expression of *ANGPTL4* (angiopoietin like 4), a *LPL* inhibitor (Bäckhed *et al.*, 2004). Thus, the increased activity of *LPL* promotes the uptake of FAs in skeletal muscle, heart, and adipose tissue, and the deposition of triglycerides in adipocytes (Scheithauer *et al.*, 2016).

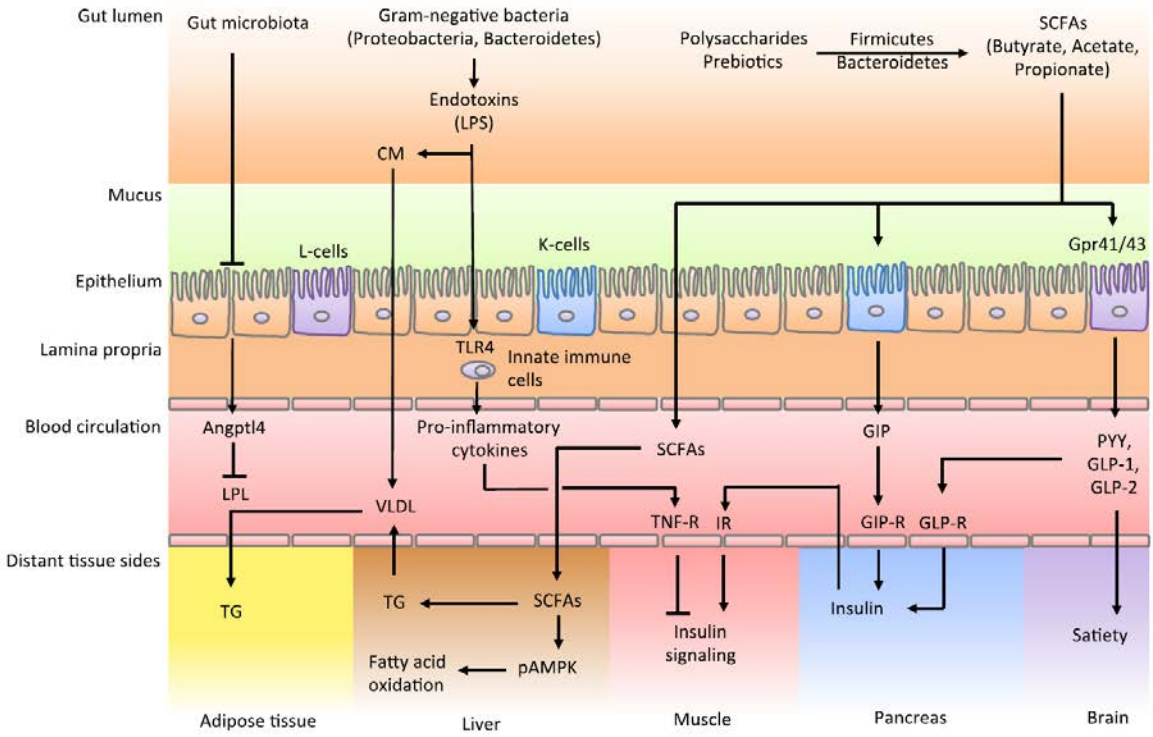


Figure 4.4. Effects of gut microbiota in weight regulation, insulin resistance and lipid metabolism (Scheithauer *et al.*, 2016).

Gut microbiota composition may affect host FA composition as well. For example, we found five regions of the pig genome (on SSC3, SSC6, SSC7, SSC9 and SSC15) that were associated with the abundance of *Akkermansia* spp. in the

rectum of 285 crossbreed pigs. In addition, *Akkermansia muciniphila* has been associated with the expression of *FASN* in adipocytes (Everard *et al.*, 2013) favouring adipose tissue adipogenesis. Indeed, *FASN* was one of the candidate genes on SSC12 associated with the abundance of myristic acid (C14:0) in the backfat of the IBMAP population, mainly in the BC1_DU animals. Further analyses are warranted in the 285 Iberian × Duroc crossbreed pigs to evaluate the association between *FASN* and the C14:0 content in backfat and to investigate if the abundance of *Akkermansia* spp. in the rectum is associated with *FASN* expression in backfat. In this scenario, individuals with a higher abundance of *Akkermansia* spp. may be producing more SCFAs that could be absorbed and affect the expression levels of *FASN*, which would elongate these SCFAs, increasing the amount of C14:0 in adipocytes.

The study of the gut microbiota has other potential applications in animal production. Certain species can be used as probiotics in substitution of antimicrobials to increase feed efficiency and health status. The improvement of feed efficiency reduces production costs, maximizing the amount of meat produced and minimizing the losses to the environment (e.g., slurry and gas emissions) (Rotz, 2004). In addition, the gut microbiota composition can be used as a measure of the health status of an animal. A healthy microbiota is often considered when it consists of a balanced microbiota composition (Liao and Nyachoti, 2017). In accordance, the reduction of the α-diversity index usually indicates individuals suffering a disease (Ott *et al.*, 2004; Drumo *et al.*, 2015; Dou *et al.*, 2017).

Antimicrobials have been systematically used for more than 60 years to improve growth and feed efficiency in productive animals (Angelakis, 2017), raising a global concern for antibiotic-resistant bacteria (Xiao *et al.*, 2016). New advances in the use of probiotics aim to reduce antimicrobial supplementation in pig fodders, which is no longer permitted in the European Union, and improve health conditions and productive traits such as weight gain and feed efficiency (Angelakis, 2017). In this manner, non-digestible food ingredients (i.e., prebiotics) are used to stimulate the growth of a reduced number of beneficial bacterial species in the gut (Gibson and Roberfroid, 1995), or these beneficial species can be provided directly as probiotics (Pandey *et al.*, 2015). Therefore, if the

microbiota composition is genetically determined it would be useful to select animals genetically associated with beneficial bacterial genera.

In the future, along with the reduction of the costs in the NGS and high-throughput technologies, the implementation of 'omics'-based tools will be more affordable for researchers. This will increment the amount of data produced and new bioinformatic pipelines and more computing power will be required. In this scenario, system biology approaches will integrate all the 'omics', including phenomics, producing a better understanding of complex traits such as meat quality and host-microbiota interactions.

5. CONCLUSIONS

1. GWAS revealed nine *Sus scrofa* chromosomal regions in SSC1, SSC2, SSC4, SSC6, SSC8, SSC10, SSC12, and SSC16 significantly associated with twelve traits for backfat fatty acid composition in three backcrosses (BC1_DU, BC1_LD and BC1_PI) with Iberian origin. A total of 33 candidate genes were proposed to explain the variations in the twelve backfat traits, including *ELOVL6*, *ELOVL7*, *FADS2* and *FASN*. In addition, six chromosomal regions in SSC4, SSC13, SSC14 and SSC17 were significantly associated with six traits for intramuscular fatty acid composition. A total of 17 candidate genes were proposed to explain the variations in the six IMF traits, including *ELOVL3* and *SCD*.

2. GWAS for the percentages of C14:0, C16:0, and C16:1(n-7) in backfat were reanalysed including seven SNPs of the *ELOVL6* gene and one SNP of the *PLA2G12A* gene, both located on SSC8. The *ELOVL6:c.-394G>A* polymorphism was the most associated SNP of SSC8 with the three fatty acids. However, *ELOVL6:c.-394G>A* was only segregating in the BC1_LD pigs.

3. GWAS for the C20:1(n-9)/C20:0 ratio in backfat was reanalysed including three SNPs of the *ELOVL7* gene and two SNPs of the *PIK3R1* gene on SSC16. The *ELOVL7:c.*1432A>G* polymorphism was the most significantly associated of the five genotyped SNPs, but other two SNPs of SSC16, rs81297480 and rs81458871, were more significant. In addition, *ELOVL7:c.*1432A>G* was not segregating in the BC1_DU pigs and no AA-genotype individuals were observed in the BC1_PI pigs. Therefore, the *ELOVL7:c.*1432A>G* SNP is unlikely to be the causal mutation of the SSC16 QTL for the C20:1(n-9)/C20:0 ratio.

4. A total of 1,928,746 indels have been identified in common from WGS data of seven IBMAP founders (two Iberian boars and five Landrace sows) using the combination of three different programs (*Dindel*, *SAMtools mpileup,* and *GATK*). Of these, 1,289 indels were predicted to have a high impact on protein sequence and function. Ten indels inside genes related

with lipid metabolism (*ASPH*, *C1QTNF12*, *CAPN9*, *CCR7*, *CRP*, *GZMA*, *JMJD1C*, *LYST*, *PEX19* and *SAMD4B*) were genotyped in animals from three different backcrosses and only the *C1QTNF12:c.557_559delCCG* polymorphism was associated with fatty acid composition, the percentage of eicosadienoic acid (C20:2(n-6)) in the intramuscular fat of BC1_PI pigs.

5. A total of 1,669 OTUs distributed in 179 genera were found in the luminal content of five gut sections (duodenum, jejunum, ileum, and proximal and distal colon) of thirteen Iberian pigs using the 16S rRNA gene sequencing method. *Lactobacillus* and *Clostridium* were the two most abundant genera in the small intestine, whereas the *Prevotella* genus was the most abundant in the colon. The colon samples were richer in species than the small intestine samples were and more similar among individuals. Metagenome predictions showed that the microbiota of the small intestine was focused on a rapid internalization and conversion of simple carbohydrates for microbial proliferation and maintenance, while a competition among anaerobic bacteria for the degradation of complex carbohydrates in the colon was found.

6. A total of 1,257 OTUs were identified in the luminal content of the rectum of 285 Iberian × Duroc pigs using the 16S rRNA gene sequencing method. These OTUs were grouped in 18 phyla and 101 genera. Firmicutes and Bacteroidetes were the two major phyla obtained. GWAS revealed 17 genomic regions of the pig genome on SSC3, SSC4, SSC6, SSC7, SSC8, SSC9, SSC10, SSC11, SSC13, SSC14, SSC15, SSC18 and SSCX associated with the relative abundance of six genera (*Akkermansia*, *CF231*, *Phascolarctobacterium*, *Prevotella*, *SMB53* and *Streptococcus*). In these genomic regions, 38 candidate genes were suggested to be modulators of the gut microbiota composition, representing a possible association between host genome and gut microbiota in pigs.

6. REFERENCES

Ahmadian, M., Wang, Y. and Sul, H. S. (2010) 'Lipolysis in adipocytes.', *The international journal of biochemistry & cell biology*, 42(5), pp. 555–9.

Albenberg, L., Esipova, T. V., Judge, C. P., Bittinger, K., Chen, J., Laughlin, A., *et al.* (2014) 'Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota', *Gastroenterology*. Elsevier, Inc, 147(5), p. 1055–1063.e8.

Albers, C. a., Lunter, G., MacArthur, D. G., McVean, G., Ouwehand, W. H. and Durbin, R. (2011) 'Dindel: Accurate indel calls from short-read data', *Genome Research*, 21(6), pp. 961–973.

Ameer, F., Scandiuzzi, L., Hasnain, S., Kalbacher, H. and Zaidi, N. (2014) 'De novo lipogenesis in health and disease.', *Metabolism: clinical and experimental*. Elsevier Inc., 63(7), pp. 895–902.

Andersson, L. and Georges, M. (2004) 'Domestic-animal genomics: deciphering the genetics of complex traits.', *Nature reviews. Genetics*, 5(3), pp. 202–12.

Angelakis, E. (2017) 'Weight gain by gut microbiota manipulation in productive animals.', *Microbial pathogenesis*. Elsevier Ltd, 106, pp. 162–170.

Archibald, A. L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M. a M., Harlizius, B., *et al.* (2010) 'Pig genome sequence - analysis and publication strategy.', *BMC genomics*, 11(1), p. 438.

Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. and Rasool, M. H. (2017) 'Proteomics: Technologies and Their Applications.', *Journal of chromatographic science*, 55(2), pp. 182–196.

Bäckhed, F., Ding, H., Wang, T., Hooper, L. V, Koh, G. Y., Nagy, A., *et al.* (2004) 'The gut microbiota as an environmental factor that regulates fat storage.', *Proceedings of the National Academy of Sciences of the United States of America*, 101(44), pp. 15718–23.

Ballester, M., Revilla, M., Puig-Oliveras, A., Marchesi, J. A. P., Castelló, A., Corominas, J., *et al.* (2016) 'Analysis of the porcine APOA2 gene expression in liver, polymorphism identification and association with fatty acid composition traits.', *Animal genetics*, 47(5), pp. 552–9.

Ballester, M., Ramayo-Caldas, Y., Revilla, M., Corominas, J., Castelló, A., Estellé, J., *et al.* (2017) 'Integration of liver gene co-expression networks and eGWAs analyses highlighted candidate regulators implicated in lipid metabolism in pigs.', *Scientific reports*. Nature Publishing Group, 7(1), p. 46539.

Barendse, W. (2014) 'Should animal fats be back on the table? A critical review of the human health effects of animal fat', *Animal Production Science*, 54(7), pp. 831–855.

Barrett, L. W., Fletcher, S. and Wilton, S. D. (2012) 'Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements.', *Cellular and molecular life sciences : CMLS*, 69(21), pp. 3613–34.

Bashan, A., Gibson, T. E., Friedman, J., Carey, V. J., Weiss, S. T., Hohmann, E. L., *et al.* (2016) 'Universality of human microbial dynamics.', *Nature*. Nature Publishing Group, 534(7606), pp. 259–62.

Benítez, R., Fernández, A., Isabel, B., Núñez, Y., De Mercado, E., Gómez-Izquierdo, E., *et al.* (2017) 'Modulatory Effects of Breed, Feeding Status, and Diet on Adipogenic, Lipogenic, and Lipolytic Gene Expression in Growing Iberian and Duroc Pigs.', *International journal of molecular sciences*, 19(1), p. 22.

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., *et al.* (2010) 'Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors', *Proceedings of the National Academy of Sciences*, 107(44), pp. 18933–18938.

Berg, J. M., Tymoczko, J. L. and Stryer, L. (2002) 'Fatty Acids Are Key Constituents of Lipids.', in *Biochemistry*. 5th edn. New York: W H Freeman.

Bergen, W. G. and Mersmann, H. J. (2005) 'Comparative aspects of lipid metabolism: impact on contemporary research and use of animal models.', *The Journal of nutrition*, 135(11), pp. 2499–502.

Bernstein, B. E., Meissner, A. and Lander, E. S. (2007) 'The mammalian epigenome.', *Cell*, 128(4), pp. 669–81.

den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J. and Bakker, B. M. (2013) 'The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism.', *Journal of lipid research*, 54(9), pp. 2325–40.

Bird, A. (2002) 'DNA methylation patterns and epigenetic memory.', *Genes & development*, 16(1), pp. 6–21.

Blasco, A. and Pena, R. N. (2018) 'Current Status of Genomic Maps: Genomic Selection/GBV in Livestock', in Niemann, H. and Wrenzycki, C. (eds) *Animal Biotechnology 2*. Cham: Springer International Publishing, p. 71.

Bloemen, J. G., Venema, K., van de Poll, M. C., Olde Damink, S. W., Buurman, W. A. and Dejong, C. H. (2009) 'Short chain fatty acids exchange across the gut and liver in humans measured at surgery.', *Clinical nutrition (Edinburgh, Scotland)*. Elsevier Ltd, 28(6), pp. 657–61.

Briolat, J., Wu, S. D., Mahata, S. K., Gonthier, B., Bagnard, D., Chasserot-Golaz, S., *et al.* (2005) 'New antimicrobial activity for the catecholamine release-inhibitory peptide from chromogranin A', *CMLS Cellular and Molecular Life Sciences*, 62(3), pp. 377–385.

Brøndum, J., Egebo, M., Agerskov, C. and Busk, H. (1998) 'On-line pork carcass grading with the Autofom ultrasound system.', *Journal of animal science*, 76(7), pp. 1859–68.

Brüssow, H. and Parkinson, S. J. (2014) 'You are what you eat', *Nature Biotechnology*, 32(3), pp. 243–245.

Bryant, M. P. (1959) 'Bacterial species of the rumen.', *Bacteriological reviews*, 23(3), pp. 125–53.

Bucher, H. C., Hengstler, P., Schindler, C. and Meier, G. (2002) 'N-3 polyunsaturated fatty acids in coronary heart disease: a meta-analysis of randomized controlled trials.', *The American journal of medicine*, 112(4), pp. 298–304.

Burrough, E. R., Arruda, B. L., Patience, J. F. and Plummer, P. J. (2015) 'Alterations in the Colonic Microbiota of Pigs Associated with Feeding Distillers Dried Grains with Solubles.', *PloS one*. Edited by M.-J. Virolle,

10(11), p. e0141337.

Camarinha-Silva, A., Maushammer, M., Wellmann, R., Vital, M., Preuss, S. and Bennewitz, J. (2017) 'Host Genome Influence on Gut Microbial Composition and Microbial Prediction of Complex Traits in Pigs.', *Genetics*, 206(3), pp. 1637–1644.

Cameron, N. . (1990) 'Genetic and phenotypic parameters for carcass traits, meat and eating quality traits in pigs', *Livestock Production Science*, 26(2), pp. 119–135.

Cameron, N. D. and Enser, M. B. (1991) 'Fatty acid composition of lipid in Longissimus dorsi muscle of Duroc and British Landrace pigs and its relationship with eating quality.', *Meat science*, 29(4), pp. 295–307.

Cao, H., Gerhold, K., Mayers, J. R., Wiest, M. M., Watkins, S. M. and Hotamisligil, G. S. (2008) 'Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism.', *Cell*, 134(6), pp. 933–44.

Chen, L., Jin, L., Li, M., Tian, S., Che, T., Tang, Q., *et al.* (2014) 'Snapshot of Structural Variations in the Tibetan Wild Boar Genome at Single-Nucleotide Resolution', *Journal of Genetics and Genomics*, 41(12), pp. 653–657.

Christie, W. W., Hunter, M. L. and Clegg, R. A. (1981) 'The effects of cerulenin on lipid metabolism in vitro in cellular preparations from the rat.', *Biochimica et biophysica acta*, 666(2), pp. 284–90.

Climent, S., Sarasa, M., Muniesa, P. and Latorre, R. (2005) 'Intestino delgado y grueso', in *Manual de anatomía y embriología de los animales domésticos: Conceptos básicos y datos aplicativos: Cabeza, aparato respiratorio, aparato digestivo y aparato urogenital*. Zaragoza: Acribia S.A.

Clop, A., Ovilo, C., Perez-Enciso, M., Cercos, A., Tomas, A., Fernandez, A., *et al.* (2003) 'Detection of QTL affecting fatty acid composition in the pig.', *Mammalian Genome*, 14(9), pp. 650–6.

Coelho, L. P., Kultima, J. R., Costea, P. I., Fournier, C., Pan, Y., Czarnecki-Maulden, G., *et al.* (2018) 'Similarity of the dog and human gut microbiomes in gene content and response to diet.', *Microbiome*.

Microbiome, 6(1), p. 72.

Cook, S. I. and Sellin, J. H. (1998) 'Review article: short chain fatty acids in health and disease.', *Alimentary pharmacology & therapeutics*, 12(6), pp. 499–507.

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Estellé, J., Castelló, A., Alves, E., *et al.* (2013a) 'Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition.', *BMC genomics*, 14, p. 843.

Corominas, J., Ramayo-Caldas, Y., Puig-Oliveras, A., Pérez-Montarelo, D., Noguera, J. L., Folch, J. M., *et al.* (2013b) 'Polymorphism in the ELOVL6 Gene Is Associated with a Major QTL Effect on Fatty Acid Composition in Pigs', *PLoS ONE*. Edited by S. Moore, 8(1), p. e53687.

Corominas, J., Marchesi, J. A. P., Puig-Oliveras, A., Revilla, M., Estellé, J., Alves, E., *et al.* (2015) 'Epigenetic regulation of the ELOVL6 gene is associated with a major QTL effect on fatty acid composition in pigs.', *Genetics, selection, evolution : GSE*, 47(1), p. 20.

Cummings, J. H. and Macfarlane, G. T. (1991) 'The control and consequences of bacterial fermentation in the human colon.', *The Journal of applied bacteriology*, 70(6), pp. 443–59.

Dettmer, K., Aronov, P. A. and Hammock, B. D. (2007) 'Mass spectrometry-based metabolomics.', *Mass spectrometry reviews*, 26(1), pp. 51–78.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014) 'Ten years of next-generation sequencing technology.', *Trends in genetics : TIG*, 30(9), pp. 418–26.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. and Thermes, C. (2018) 'The Third Revolution in Sequencing Technology.', *Trends in genetics : TIG*. Elsevier Ltd, 34(9), pp. 666–681.

Donaldson, G. P., Lee, S. M. and Mazmanian, S. K. (2016) 'Gut biogeography of the bacterial microbiota.', *Nature reviews. Microbiology*, 14(1), pp. 20–32.

Dou, S., Gadonna-Widehem, P., Rome, V., Hamoudi, D., Rhazi, L., Lakhal, L., *et*

al. (2017) 'Characterisation of Early-Life Fecal Microbiota in Susceptible and Healthy Pigs to Post-Weaning Diarrhoea', *PloS one*, 12(1), p. e0169851.

Drumo, R., Pesciaroli, M., Ruggeri, J., Tarantino, M., Chirullo, B., Pistoia, C., *et al.* (2015) 'Salmonella enterica Serovar Typhimurium Exploits Inflammation to Modify Swine Intestinal Microbiota.', *Frontiers in cellular and infection microbiology*, 5(January), p. 106.

Eaton, S., Bartlett, K. and Pourfarzam, M. (1996) 'Mammalian mitochondrial beta-oxidation.', *The Biochemical journal*, 320 (Pt 2(2), pp. 345–57.

Eichenbaum, Z., Muller, E., Morse, S. A. and Scott, J. R. (1996) 'Acquisition of iron from host proteins by the group A streptococcus.', *Infection and immunity*, 64(12), pp. 5428–9.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., *et al.* (2009) 'Real-time DNA sequencing from single polymerase molecules.', *Science (New York, N.Y.)*, 323(5910), pp. 133–8.

Eng, J. K., McCormack, A. L. and Yates, J. R. (1994) 'An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.', *Journal of the American Society for Mass Spectrometry*, 5(11), pp. 976–89.

Enomoto, T., Ohashi, K., Shibata, R., Higuchi, A., Maruyama, S., Izumiya, Y., *et al.* (2011) 'Adipolin/C1qdc2/CTRP12 protein functions as an adipokine that improves glucose metabolism.', *The Journal of biological chemistry*, 286(40), pp. 34552–8.

Estany, J., Ros-Freixedes, R., Tor, M. and Pena, R. N. (2014) 'A functional variant in the stearoyl-CoA desaturase gene promoter enhances fatty acid desaturation in pork.', *PloS one*. Edited by J.-M. A. Lobaccaro, 9(1), p. e86177.

Estellé, J., Mercadé, A., Noguera, J. L., Pérez-Enciso, M., Ovilo, C., Sánchez, A., *et al.* (2005) 'Effect of the porcine IGF2-intron3-G3072A substitution in an outbred Large White population and in an Iberian x Landrace cross.', *Journal of animal science*, 83(12), pp. 2723–8.

Estellé, J., Pérez-Enciso, M., Mercadé, A., Varona, L., Alves, E., Sánchez, A., *et al.* (2006) 'Characterization of the porcine FABP5 gene and its association with the FAT1 QTL in an Iberian by Landrace cross.', *Animal genetics*, 37(6), pp. 589–91.

Estellé, J., Mach, N., Ramayo-Caldas, Y., Levenez, F., Lemonnier, G., Denis, C., *et al.* (2014) 'The influence of host's genetics on the gut microbiota composition in pigs and its links with immunity traits', in *10th World Congress of Genetics Applied to Livestock Production, Vancouver, BC, Canada*.

van Eunen, K., Volker-Touw, C. M. L., Gerding, A., Bleeker, A., Wolters, J. C., van Rijt, W. J., *et al.* (2016) 'Living on the edge: substrate competition explains loss of robustness in mitochondrial fatty-acid oxidation disorders.', *BMC biology*. BMC Biology, 14(1), p. 107.

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J. P., Druart, C., Bindels, L. B., *et al.* (2013) 'Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(22), pp. 9066–71.

Fagarasan, S., Muramatsu, M., Suzuki, K., Nagaoka, H., Hiai, H. and Honjo, T. (2002) 'Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora.', *Science (New York, N.Y.)*, 298(5597), pp. 1424–7.

Fernández, A. I., Pérez-Montarelo, D., Barragán, C., Ramayo-Caldas, Y., Ibañez-Escriche, N., Castelló, A., *et al.* (2012) 'Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip', *BMC Genetics*, p. 41.

Fisher, R. A. (1918) 'The Correlation between Relatives on the Supposition of Mendelian Inheritance.', *Transactions of the Royal Society of Edinburgh*, 52, pp. 399–433.

Frühbeck, G., Méndez-Giménez, L., Fernández-Formoso, J.-A., Fernández, S. and Rodríguez, A. (2014) 'Regulation of adipocyte lipolysis.', *Nutrition*

research reviews, 27(1), pp. 63–93.

Fujii, J., Otsu, K., Zorzato, F., de Leon, S., Khanna, V. K., Weiler, J. E., *et al.* (1991) 'Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia.', *Science (New York, N.Y.)*, 253(5018), pp. 448–51.

Gavino, G. R. and Gavino, V. C. (1991) 'Rat liver outer mitochondrial carnitine palmitoyltransferase activity towards long-chain polyunsaturated fatty acids and their CoA esters.', *Lipids*, 26(4), pp. 266–70.

Gibson, G. R. and Roberfroid, M. B. (1995) 'Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics.', *The Journal of nutrition*, 125(6), pp. 1401–12.

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. and Field, K. G. (1990) 'Genetic diversity in Sargasso Sea bacterioplankton.', *Nature*, 345(6270), pp. 60–3.

Goddard, M. E. and Hayes, B. J. (2009) 'Mapping genes for complex traits in domestic animals and their use in breeding programmes.', *Nature reviews. Genetics*. Nature Publishing Group, 10(6), pp. 381–91.

Goddard, M. E. and Whitelaw, E. (2014) 'The use of epigenetic phenomena for the improvement of sheep and cattle.', *Frontiers in genetics*, 5(JUL), p. 247.

Gol, S., Pena, R. N., Rothschild, M. F., Tor, M. and Estany, J. (2018) 'A polymorphism in the fatty acid desaturase-2 gene is associated with the arachidonic acid metabolism in pigs.', *Scientific reports*. Springer US, 8(1), p. 14336.

Goll, M. G. and Bestor, T. H. (2005) 'Eukaryotic cytosine methyltransferases.', *Annual review of biochemistry*, 74(1), pp. 481–514.

Gonzalez, L. M., Moeser, A. J. and Blikslager, A. T. (2015) 'Porcine models of digestive disease: the future of large animal translational research.', *Translational research : the journal of laboratory and clinical medicine*, 166(1), pp. 12–27.

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., *et al.* (2014) 'Human genetics shape the gut microbiome.', *Cell*. Elsevier Inc., 159(4), pp. 789–99.

Gorvitovskaia, A., Holmes, S. P. and Huse, S. M. (2016) 'Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle.', *Microbiome*. Microbiome, 4(1), p. 15.

Gray, M. W., Sankoff, D. and Cedergren, R. J. (1984) 'On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA.', *Nucleic acids research*, 12(14), pp. 5837–52.

Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., *et al.* (2012) 'Analyses of pig genomes provide insight into porcine demography and evolution.', *Nature*. Nature Publishing Group, 491(7424), pp. 393–8.

Groenen, M. A. M. (2015) 'Development of a high-density Axiom® porcine genotyping array to meet research and commercial needs', in *Plant and Animal Genome XXIII Conference, San Diego, CA, USA*.

Groenen, M. A. M. (2016) 'A decade of pig genome sequencing: a window on pig domestication and evolution.', *Genetics, selection, evolution : GSE*. BioMed Central, 48(1), p. 23.

Guillou, H., D'andrea, S., Rioux, V., Jan, S. and Legrand, P. (2004) 'The surprising diversity of Delta6-desaturase substrates.', *Biochemical Society transactions*, 32(Pt 1), pp. 86–7.

Guillou, H., Zadravec, D., Martin, P. G. P. and Jacobsson, A. (2010) 'The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice.', *Progress in lipid research*. Elsevier Ltd, 49(2), pp. 186–99.

Guo, F., Ju, F., Cai, L. and Zhang, T. (2013) 'Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment.', *PloS one*. Edited by A. O. Hudson, 8(10), p. e76185.

Haley, C. S., Archibald, A., Anderson, L., Bosma, A. A., Davies, W., Fredholm, M., *et al.* (1990) 'The pig gene mapping project – PiGMaP.', in *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production*, pp. 67–70.

Hayes, B. and Goddard, M. E. (2001) 'The distribution of the effects of genes affecting quantitative traits in livestock.', *Genetics, selection, evolution : GSE*, 33(3), pp. 209–29.

He, M., Fang, S., Huang, X., Zhao, Y., Ke, S., Yang, H., *et al.* (2016) 'Evaluating the Contribution of Gut Microbiota to the Variation of Porcine Fatness with the Cecum and Fecal Samples', *Frontiers in Microbiology*, 07, p. 2108.

Hiraoka, T., Fukuwatari, T., Imaizumi, M. and Fushiki, T. (2003) 'Effects of oral stimulation with fats on the cephalic phase of pancreatic enzyme secretion in esophagostomized rats.', *Physiology & behavior*, 79(4–5), pp. 713–7.

Holman, D. B., Brunelle, B. W., Trachsel, J. and Allen, H. K. (2017) 'Meta-analysis To Define a Core Microbiota in the Swine Gut.', *mSystems*. Edited by H. Bik, 2(3), pp. e00004-17.

Houle, D., Govindaraju, D. R. and Omholt, S. (2010) 'Phenomics: the next challenge.', *Nature reviews. Genetics*. Nature Publishing Group, 11(12), pp. 855–66.

Høverstad, T. and Midtvedt, T. (1986) 'Short-chain fatty acids in germfree mice and rats.', *The Journal of nutrition*, 116(9), pp. 1772–6.

Hu, Z.-L., Park, C. A. and Reecy, J. M. (2016) 'Developmental progress and current status of the Animal QTLdb.', *Nucleic acids research*, 44(D1), pp. D827-33.

Hughes, S. A., Shewry, P. R., Gibson, G. R., McCleary, B. V. and Rastall, R. A. (2008) 'In vitro fermentation of oat and barley derived beta-glucans by human faecal microbiota.', *FEMS microbiology ecology*, 64(3), pp. 482–93.

Humphray, S. J., Scott, C. E., Clark, R., Marron, B., Bender, C., Camm, N., *et al.* (2007) 'A high utility integrated map of the pig genome.', *Genome biology*, 8(7), p. R139.

International Human Genome Sequencing Consortium (2004) 'Finishing the euchromatic sequence of the human genome.', *Nature*, 431(7011), pp. 931–45.

Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B. and Akeson, M. (2015) 'Improved data analysis for the MinION nanopore sequencer.', *Nature methods*, 12(4), pp. 351–6.

Jiménez-Colmenero, F., Ventanas, J. and Toldrá, F. (2010) 'Nutritional composition of dry-cured ham and its role in a healthy diet.', *Meat science*. Elsevier Ltd, 84(4), pp. 585–93.

Kang, H., Wang, H., Fan, Z., Zhao, P., Khan, A., Yin, Z., *et al.* (2015) 'Resequencing diverse Chinese indigenous breeds to enrich the map of genomic variations in swine.', *Genomics*. Elsevier Inc., 106(5), pp. 286–94.

Katouli, M., Lund, A., Wallgren, P., Kühn, I., Söderlind, O. and Möllby, R. (1997) 'Metabolic fingerprinting and fermentative capacity of the intestinal flora of pigs during pre- and post-weaning periods.', *Journal of applied microbiology*, 83(2), pp. 147–54.

Kelly, J., Daly, K., Moran, A. W., Ryan, S., Bravo, D. and Shirazi-Beechey, S. P. (2017) 'Composition and diversity of mucosa-associated microbiota along the entire length of the pig gastrointestinal tract; dietary influences.', *Environmental microbiology*, 19(4), pp. 1425–1438.

Kers, J. G., Velkers, F. C., Fischer, E. A. J., Hermes, G. D. A., Stegeman, J. A. and Smidt, H. (2018) 'Host and environmental factors affecting the intestinal microbiota in chickens', *Frontiers in Microbiology*, 9(FEB), pp. 1–14.

Kim, H. B. and Isaacson, R. E. (2017) 'Salmonella in Swine: Microbiota Interactions.', *Annual review of animal biosciences*, 5(1), pp. 43–63.

Korte, A. and Farlow, A. (2013) 'The advantages and limitations of trait analysis with GWAS: a review.', *Plant methods*. Plant Methods, 9(1), p. 29.

Kouba, M. and Sellier, P. (2011) 'A review of the factors influencing the development of intermuscular adipose tissue in the growing pig.', *Meat*

science. Elsevier Ltd, 88(2), pp. 213–20.

Van Laere, A.-S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., *et al.* (2003) 'A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig.', *Nature*, 425(6960), pp. 832–6.

Lagarde, M., Bernoud-Hubac, N., Calzada, C., Véricel, E. and Guichardant, M. (2013) 'Lipidomics of essential fatty acids and oxygenated metabolites.', *Molecular nutrition & food research*, 57(8), pp. 1347–58.

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., *et al.* (2005) 'Worldwide phylogeography of wild boar reveals multiple centers of pig domestication.', *Science (New York, N.Y.)*, 307(5715), pp. 1618–21.

Lederberg, J. and McCray, A. (2001) ''Ome Sweet 'Omics—a genealogical treasury of words', *Scientist*, 15, p. 8.

Legrand, P. and Rioux, V. (2010) 'The complex and important cellular and metabolic functions of saturated fatty acids.', *Lipids*, 45(10), pp. 941–6.

Ley, R., Turnbaugh, P., Klein, S. and Gordon, J. (2006) 'Microbial ecology: human gut microbes associated with obesity.', *Nature*, 444(7122), pp. 1022–3.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079.

Liao, S. F. and Nyachoti, M. (2017) 'Using probiotics to improve swine gut health and nutrient utilization.', *Animal nutrition*. Elsevier Taiwan LLC, 3(4), pp. 331–343.

Lin, J., Zhang, S. M., Cook, N. R., Lee, I.-M. and Buring, J. E. (2004) 'Dietary fat and fatty acids and risk of colorectal cancer in women.', *American journal of epidemiology*, 160(10), pp. 1011–22.

Liu, H., Ivarsson, E., Dicksved, J., Lundh, T. and Lindberg, J. E. (2012) 'Inclusion of Chicory (Cichorium intybus L.) in pigs' diets affects the intestinal microenvironment and the gut microbiota', *Applied and Environmental Microbiology*, 78(12), pp. 4102–4109.

Liu, Y., Beyer, A. and Aebersold, R. (2016) 'On the Dependency of Cellular

Protein Levels on mRNA Abundance.', *Cell*. Elsevier Inc., 165(3), pp. 535–50.

Liu, Z., DeSantis, T. Z., Andersen, G. L. and Knight, R. (2008) 'Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers.', *Nucleic acids research*, 36(18), p. e120.

Looft, T., Johnson, T. A., Allen, H. K., Bayles, D. O., Alt, D. P., Stedtfeld, R. D., *et al.* (2012) 'In-feed antibiotic effects on the swine intestinal microbiome.', *Proceedings of the National Academy of Sciences of the United States of America*, 109(5), pp. 1691–6.

Louis, P., Scott, K. P., Duncan, S. H. and Flint, H. J. (2007) 'Understanding the effects of diet on bacterial metabolism in the large intestine.', *Journal of applied microbiology*, 102(5), pp. 1197–208.

Lu, D., Tiezzi, F., Schillebeeckx, C., McNulty, N. P., Schwab, C., Shull, C., *et al.* (2018) 'Host contributes to longitudinal diversity of fecal microbiota in swine selected for lean growth.', *Microbiome*. Microbiome, 6(1), p. 4.

Lupski, J. R., Roth, J. R. and Weinstock, G. M. (1996) 'Chromosomal duplications in bacteria, fruit flies, and humans.', *American journal of human genetics*, 58(1), pp. 21–7.

Mach, N., Berri, M., Estellé, J., Levenez, F., Lemonnier, G., Denis, C., *et al.* (2015) 'Early-life establishment of the swine gut microbiome and impact on host phenotypes', *Environmental Microbiology Reports*, 7(3), pp. 554–569.

Mackay, T. F. (2001) 'Quantitative trait loci in Drosophila.', *Nature reviews. Genetics*, 2(1), pp. 11–20.

Mackay, T. F. C., Stone, E. A. and Ayroles, J. F. (2009) 'The genetics of quantitative traits: challenges and prospects.', *Nature reviews. Genetics*, 10(8), pp. 565–77.

Maltecca, C., Lu, D., Schillebeeckx, C., McNulty, N., Schawb, C., Schull, C., *et al.* (2018) 'Predicting Growth and Carcass Traits in Swine Using Metagenomic Data and Machine Learning Algorithms', *bioRxiv*.

Marchesi, J. R. and Ravel, J. (2015) 'The vocabulary of microbiome research: a proposal.', *Microbiome*. Microbiome, 3(1), p. 31.

Mardis, E. R. (2008) 'The impact of next-generation sequencing technology on genetics', *Trends in Genetics*, 24(3), pp. 133–141.

Margueron, R., Trojer, P. and Reinberg, D. (2005) 'The key to development: interpreting the histone code?', *Current opinion in genetics & development*, 15(2), pp. 163–76.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., *et al.* (2005) 'Genome sequencing in microfabricated high-density picolitre reactors.', *Nature*, 437(7057), pp. 376–80.

Marteyn, B., Scorza, F. B., Sansonetti, P. J. and Tang, C. (2011) 'Breathing life into pathogens: the influence of oxygen on bacterial virulence and host responses in the gastrointestinal tract.', *Cellular microbiology*, 13(2), pp. 171–6.

Matsuzaka, T. and Shimano, H. (2009) 'Elovl6: a new player in fatty acid metabolism and insulin sensitivity.', *Journal of molecular medicine (Berlin, Germany)*, 87(4), pp. 379–84.

Mattson, F. H. and Grundy, S. M. (1985) 'Comparison of effects of dietary saturated, monounsaturated, and polyunsaturated fatty acids on plasma lipids and lipoproteins in man.', *Journal of lipid research*, 26(2), pp. 194–202.

McIntyre, M. K., Peacock, T. J., Akers, K. S. and Burmeister, D. M. (2016) 'Initial Characterization of the Pig Skin Bacteriome and Its Effect on In Vitro Models of Wound Healing.', *PloS one*. Edited by J. M. Brandner, 11(11), p. e0166176.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* (2010) 'The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, 20(9), pp. 1297–1303.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) 'Deriving the consequences of genomic variants with the Ensembl

API and SNP Effect Predictor', *Bioinformatics*, 26(16), pp. 2069–2070.

Mercadé, A., Estellé, J., Noguera, J. L., Folch, J. M., Varona, L., Silió, L., *et al.* (2005) 'On growth, fatness, and form: a further look at porcine chromosome 4 in an Iberian x Landrace cross.', *Mammalian genome : official journal of the International Mammalian Genome Society*, 16(5), pp. 374–82.

Mercadé, A., Pérez-Enciso, M., Varona, L., Alves, E., Noguera, J. L., Sánchez, A., *et al.* (2006) 'Adipocyte fatty-acid binding protein is closely associated to the porcine FAT1 locus on chromosome 4.', *Journal of animal science*, 84(11), pp. 2907–13.

Meuwissen, T. H. E. and Goddard, M. E. (1996) 'The use of marker haplotypes in animal breeding schemes', *Genetics Selection Evolution*, 28(2), pp. 161–176.

Meuwissen, T. H. and Goddard, M. E. (2000) 'Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci.', *Genetics*, 155(1), pp. 421–30.

Meuwissen, T. H., Hayes, B. J. and Goddard, M. E. (2001) 'Prediction of total genetic value using genome-wide dense marker maps.', *Genetics*, 157(4), pp. 1819–29.

Milan, D., Jeon, J. T., Looft, C., Amarger, V., Robic, A., Thelander, M., *et al.* (2000) 'A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle.', *Science (New York, N.Y.)*, 288(5469), pp. 1248–51.

Miles, C. and Wayne, M. (2008) 'Quantitative trait locus (QTL) analysis.', *Nature Education*, 1(1), p. 208.

Million, M., Maraninchi, M., Henry, M., Armougom, F., Richet, H., Carrieri, P., *et al.* (2012) 'Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii.', *International journal of obesity (2005)*. Nature Publishing Group, 36(6), pp. 817–25.

Mills, R. E., Pittard, W. S., Mullaney, J. M., Farooq, U., Creasy, T. H., Mahurkar,

A. a., *et al.* (2011) 'Natural genetic variation caused by small insertions and deletions in the human genome', *Genome Research*, 21(6), pp. 830–839.

Molnár, J., Nagy, T., Stéger, V., Tóth, G., Marincs, F. and Barta, E. (2014) 'Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary.', *BMC genomics*, 15(761), p. 761.

Muñoz, G., Alves, E., Fernández, A., Ovilo, C., Barragán, C., Estellé, J., *et al.* (2007) 'QTL detection on porcine chromosome 12 for fatty-acid composition and association analyses of the fatty acid synthase, gastric inhibitory polypeptide and acetyl-coenzyme A carboxylase alpha genes.', *Animal genetics*, 38(6), pp. 639–46.

Muñoz, M., Rodríguez, M. C., Alves, E., Folch, J. M., Ibañez-Escriche, N., Silió, L., *et al.* (2013) 'Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data', *BMC Genomics*, 14(1), p. 845.

Murphy, E. F., Cotter, P. D., Healy, S., Marques, T. M., O'Sullivan, O., Fouhy, F., *et al.* (2010) 'Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models.', *Gut*, 59(12), pp. 1635–42.

Naganuma, T., Sato, Y., Sassa, T., Ohno, Y. and Kihara, A. (2011) 'Biochemical characterization of the very long-chain fatty acid elongase ELOVL7.', *FEBS letters*. Federation of European Biochemical Societies, 585(20), pp. 3337–41.

National Institutes of Health (2018) *Talking Glossary of Genetic Terms.* Retrieved from National Human Genome Research Institute: https://www.genome.gov/glossary/.

Neethling, J., Hoffman, L. C. and Muller, M. (2016) 'Factors influencing the flavour of game meat: A review.', *Meat science*. Elsevier Ltd, 113, pp. 139–53.

NIH HMP Working Group, Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., *et al.* (2009) 'The NIH Human Microbiome Project.', *Genome research*, 19(12), pp. 2317–23.

Nussbaum, R. L., McInnes, R. R. and Willard, H. F. (2008) *Genética en medicina*. 7ª. Philadelphia: Elsevier Masson, Saunders.

O'Hea, E. K. and Leveille, G. A. (1968) 'Lipid metabolism in isolated adipose tissue of the domestic pig (Sus domesticus).', *Comparative biochemistry and physiology*, 26(3), pp. 1081–9.

O'Hea, E. K. and Leveille, G. A. (1969) 'Significance of adipose tissue and liver as sites of fatty acid synthesis in the pig and the efficiency of utilization of various substrates for lipogenesis.', *The Journal of nutrition*, 99(3), pp. 338–44.

Ohashi, H., Hasegawa, M., Wakimoto, K. and Miyamoto-Sato, E. (2015) 'Next-generation technologies for multiomics approaches including interactome sequencing.', *BioMed research international*, 2015(Table 2), p. 104209.

Ojeda, A., Rozas, J., Folch, J. M. and Pérez-Enciso, M. (2006) 'Unexpected high polymorphism at the FABP4 gene unveils a complex history for pig populations.', *Genetics*, 174(4), pp. 2119–27.

Oldenbroek, K. and van der Waaij, L. (2014) 'Origin of animal breeding: a history of science.', in *Textbook animal breeding: Animal breeding and genetics for BSc students.* Centre for Genetic Resources and Animal Breeding and Genomics Group, Wageningen University and Research Centre, the Netherlands: Groen Kennisnet, pp. 21–22.

Osmundsen, H. and Bjørnstad, K. (1985) 'Inhibitory effects of some long-chain unsaturated fatty acids on mitochondrial beta-oxidation. Effects of streptozotocin-induced diabetes on mitochondrial beta-oxidation of polyunsaturated fatty acids.', *The Biochemical journal*, 230(2), pp. 329–37.

Ott, S. J., Musfeldt, M., Wenderoth, D. F., Hampe, J., Brant, O., Fölsch, U. R., *et al.* (2004) 'Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease.', *Gut*, 53(5), pp. 685–93.

Ottman, N., Geerlings, S. Y., Aalvink, S., de Vos, W. M. and Belzer, C. (2017) 'Action and function of Akkermansia muciniphila in microbiome ecology, health and disease.', *Best practice & research. Clinical gastroenterology*.

Elsevier Ltd, 31(6), pp. 637–642.

Ovilo, C., Pérez-Enciso, M., Barragán, C., Clop, A., Rodríquez, C., Oliver, M. A., *et al.* (2000) 'A QTL for intramuscular fat and backfat thickness is located on porcine chromosome 6.', *Mammalian genome : official journal of the International Mammalian Genome Society*, 11(4), pp. 344–6.

Ovilo, C., Clop, A., Noguera, J. L., Oliver, M. A., Barragán, C., Rodriguez, C., *et al.* (2002) 'Quantitative trait locus mapping for meat quality traits in an Iberian x Landrace F2 pig population.', *Journal of animal science*, 80(11), pp. 2801–8.

Ovilo, C., Fernández, A., Noguera, J. L., Barragán, C., Letón, R., Rodríguez, C., *et al.* (2005) 'Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross.', *Genetical research*, 85(1), pp. 57–67.

Palsson, B. (2002) 'In silico biology through "omics".', *Nature biotechnology*, 20(7), pp. 649–50.

Pandey, V., Berwal, V., Solanki, N. and Malik, N. S. (2015) 'Probiotics: Healthy bugs and nourishing elements of diet.', *Journal of International Society of Preventive & Community Dentistry*, 5(2), pp. 81–7.

Patti, G. J., Yanes, O. and Siuzdak, G. (2012) 'Innovation: Metabolomics: the apogee of the omics trilogy.', *Nature reviews. Molecular cell biology*. Nature Publishing Group, 13(4), pp. 263–9.

Pedersen, R., Andersen, A. D., Mølbak, L., Stagsted, J. and Boye, M. (2013) 'Changes in the gut microbiota of cloned and non-cloned control pigs during development of obesity: gut microbiota during development of obesity in cloned pigs.', *BMC microbiology*, 13(1), p. 30.

Pellegrini, M. and Ferrari, R. (2012) 'Epigenetic Analysis: ChIP-chip and ChIP-seq', in *Methods in molecular biology (Clifton, N.J.)*, pp. 377–387.

Pereira, F. C. and Berry, D. (2017) 'Microbial nutrient niches in the gut.', *Environmental microbiology*, 19(4), pp. 1366–1378.

Pérez-Enciso, M., Clop, A., Noguera, J. L., Ovilo, C., Coll, A., Folch, J. M., *et al.*

(2000) 'A QTL on pig chromosome 4 affects fatty acid metabolism: evidence from an Iberian by Landrace intercross.', *Journal of animal science*, 78(10), pp. 2525–31.

Pérez-Enciso, M., Clop, A., Folch, J. M., Sánchez, A., Oliver, M. A., Ovilo, C., *et al.* (2002) 'Exploring alternative models for sex-linked quantitative trait loci in outbred populations: application to an iberian x landrace pig intercross.', *Genetics*, 161(4), pp. 1625–32.

Pérez-Enciso, M., Mercadé, A., Bidanel, J. P., Geldermann, H., Cepica, S., Bartenschlager, H., *et al.* (2005) 'Large-scale, multibreed, multitrait analyses of quantitative trait loci experiments: the case of porcine X chromosome.', *Journal of animal science*, 83(10), pp. 2289–96.

Pérez-Enciso, M., Rincón, J. C. and Legarra, A. (2015) 'Sequence- vs. chip-assisted genomic selection: accurate biological information is advised.', *Genetics, selection, evolution : GSE*, 47(1), p. 43.

Puig-Oliveras, A., Ramayo-Caldas, Y., Corominas, J., Estellé, J., Pérez-Montarelo, D., Hudson, N. J., *et al.* (2014) 'Differences in Muscle Transcriptome among Pigs Phenotypically Extreme for Fatty Acid Composition.', *PloS one*, 9(6), p. e99720.

Puig-Oliveras, A., Revilla, M., Castelló, A., Fernández, A. I., Folch, J. M. and Ballester, M. (2016) 'Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat.', *Scientific reports*. Nature Publishing Group, 6(1), p. 31803.

Quan, J., Cai, G., Ye, J., Yang, M., Ding, R., Wang, X., *et al.* (2018) 'A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios.', *Scientific reports*. Springer US, 8(1), p. 4536.

Ramayo-Caldas, Y., Mercadé, A., Castelló, A., Yang, B., Rodríguez, C., Alves, E., *et al.* (2012a) 'Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross.', *Journal of animal science*, 90(9), pp. 2883–93.

Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A., Corominas, J., Castelló, A., Ballester, M., *et al.* (2012b) 'Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition.', *BMC genomics*, 13(1), p. 547.

Ramos, A. M., Crooijmans, R. P. M. a, Affara, N. a, Amaral, A. J., Archibald, A. L., Beever, J. E., *et al.* (2009) 'Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.', *PloS one*. Edited by L. Orban, 4(8), p. e6524.

Ren, Z., Liu, W., Zheng, R., Zuo, B., Xu, D., Lei, M., *et al.* (2012) 'A 304 bp insertion/deletion mutation in promoter region induces the increase of porcine IDH3β gene expression.', *Molecular biology reports*, 39(2), pp. 1419–26.

Revilla, M., Ramayo-Caldas, Y., Castelló, A., Corominas, J., Puig-Oliveras, A., Ibáñez-Escriche, N., *et al.* (2014) 'New insight into the SSC8 genetic determination of fatty acid composition in pigs.', *Genetics, selection, evolution : GSE*, 46(1), p. 28.

Revilla, M., Puig-Oliveras, A., Castelló, A., Crespo-Piazuelo, D., Paludo, E., Fernández, A. I., *et al.* (2017) 'A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits.', *PloS one*. Edited by R. Davoli, 12(5), p. e0177014.

Revilla, M., Puig-Oliveras, A., Crespo-Piazuelo, D., Criado-Mesas, L., Castelló, A., Fernández, A. I., *et al.* (2018) 'Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions.', *Scientific reports*, 8(1), p. 2045.

Rioux, V., Choque, B., Ezanno, H., Duby, C., Catheline, D. and Legrand, P. (2015) 'Influence of the cis-9, cis-12 and cis-15 double bond position in octadecenoic acid (18:1) isomers on the rat FADS2-catalyzed Δ6-desaturation.', *Chemistry and physics of lipids*. Elsevier Ireland Ltd, 187, pp. 10–9.

Rioux, V., Catheline, D. and Legrand, P. (2007) 'In rat hepatocytes, myristic acid occurs through lipogenesis, palmitic acid shortening and lauric acid elongation.', *Animal : an international journal of animal bioscience*, 1(6), pp. 820–6.

Ros-Freixedes, R., Gol, S., Pena, R. N., Tor, M., Ibáñez-Escriche, N., Dekkers, J. C. M., *et al.* (2016) 'Genome-Wide Association Study Singles Out SCD and LEPR as the Two Main Loci Influencing Intramuscular Fat Content and Fatty Acid Composition in Duroc Pigs.', *PloS one*, 11(3), p. e0152496.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., *et al.* (2018) 'Environment dominates over host genetics in shaping human gut microbiota.', *Nature*. Nature Publishing Group, 555(7695), pp. 210–215.

Rotz, C. A. (2004) 'Management to reduce nitrogen losses in animal production.', *Journal of animal science*, 82 E-Suppl(3), pp. E119-137.

Samorè, A. B. and Fontanesi, L. (2016) 'Genomic selection in pigs: state of the art and perspectives', *Italian Journal of Animal Science*, 15(2), pp. 211–232.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors.', *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp. 5463–7.

Sañudo Astiz, C. (2008) *Manual de diferenciación racial.* Zaragoza: Servet, Diseño y comunicación S.L.

Sato, S., Uemoto, Y., Kikuchi, T., Egawa, S., Kohira, K., Saito, T., *et al.* (2017) 'Genome-wide association studies reveal additional related loci for fatty acid composition in a Duroc pig multigenerational population.', *Animal science journal*, 88(10), pp. 1482–1490.

Schadt, E. E., Turner, S. and Kasarskis, A. (2010) 'A window into third-generation sequencing.', *Human molecular genetics*, 19(R2), pp. R227-40.

Scheithauer, T. P. M., Dallinga-Thie, G. M., de Vos, W. M., Nieuwdorp, M. and van Raalte, D. H. (2016) 'Causality of small and large intestinal microbiota in weight regulation and insulin resistance.', *Molecular metabolism*.

Elsevier GmbH, 5(9), pp. 759–70.

Schook, L. B., Beever, J. E., Rogers, J., Humphray, S., Archibald, A., Chardon, P., *et al.* (2005) 'Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome.', *Comparative and functional genomics*, 6(4), pp. 251–5.

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N. A., Donus, C., *et al.* (2010) 'Microbiota and SCFA in lean and overweight healthy subjects.', *Obesity (Silver Spring, Md.)*. Nature Publishing Group, 18(1), pp. 190–5.

Serra, X., Gil, F., Pérez-Enciso, M., Oliver, M. ., Vázquez, J. ., Gispert, M., *et al.* (1998) 'A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyerbas line) and Landrace pigs', *Livestock Production Science*, 56(3), pp. 215–223.

Sharpe, A. J. and McKenzie, M. (2018) 'Mitochondrial Fatty Acid Oxidation Disorders Associated with Short-Chain Enoyl-CoA Hydratase (ECHS1) Deficiency.', *Cells*, 7(6), p. 46.

Slifierz, M. J., Friendship, R. M. and Weese, J. S. (2015) 'Longitudinal study of the early-life fecal and nasal microbiotas of the domestic pig.', *BMC microbiology*. BMC Microbiology, 15(1), p. 184.

van Son, M., Enger, E. G., Grove, H., Ros-Freixedes, R., Kent, M. P., Lien, S., *et al.* (2017) 'Genome-wide association study confirm major QTL for backfat fatty acid composition on SSC14 in Duroc pigs.', *BMC genomics*. BMC Genomics, 18(1), p. 369.

Spor, A., Koren, O. and Ley, R. (2011) 'Unravelling the effects of the environment and host genotype on the gut microbiome.', *Nature reviews. Microbiology*. Nature Publishing Group, 9(4), pp. 279–90.

Stock, J., Calderón Díaz, J., Abell, C., Baas, T., Rothschild, M. F., Mote, B. E., *et al.* (2017) 'Development of an Objective Feet and Leg Conformation Evaluation Method Using Digital Imagery in Swine.', *Journal of Animal Sciences and Livestock Production*, 1, pp. 1–7.

Su, Y., Bian, G., Zhu, Z., Smidt, H. and Zhu, W. (2014) 'Early methanogenic colonisation in the faeces of Meishan and Yorkshire piglets as determined

by pyrosequencing analysis.', *Archaea (Vancouver, B.C.)*. Hindawi Publishing Corporation, 2014, p. 547908.

Subramaniam, S., Fahy, E., Gupta, S., Sud, M., Byrnes, R. W., Cotter, D., *et al.* (2011) 'Bioinformatics and systems biology of the lipidome.', *Chemical reviews*, 111(10), pp. 6452–90.

Tazzini, N. (2013) *Long chain fatty acid synthesis*. Retrieved from http://www.tuscany-diet.net/2013/02/03/long-chain-fatty-acid-synthesis/.

Thompson, C. L., Wang, B. and Holmes, A. J. (2008) 'The immediate environment during postnatal development has long-term impact on gut community structure in pigs.', *The ISME journal*, 2(7), pp. 739–48.

Tibau, J. (1992) 'Mejora genética de los caracteres productivos en el ganado porcino', *Torcí*, 7, pp. 11–21.

Tsai, S. F., Bishop, D. F. and Desnick, R. J. (1988) 'Human uroporphyrinogen III synthase: molecular cloning, nucleotide sequence, and expression of a full-length cDNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 85(19), pp. 7049–53.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. a, Magrini, V., Mardis, E. R. and Gordon, J. I. (2006) 'An obesity-associated gut microbiome with increased capacity for energy harvest.', *Nature*, 444(7122), pp. 1027–31.

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M. S., Kevans, D., Smith, M. I., *et al.* (2016) 'Association of host genome with intestinal microbial composition in a large healthy cohort.', *Nature genetics*, 48(11), pp. 1413–1417.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., *et al.* (2004) 'Community structure and metabolism through reconstruction of microbial genomes from the environment.', *Nature*, 428(6978), pp. 37–43.

U.S. National Library of Medicine (2018) *Help Me Understand Genetics: Mutations and Health*. Retrieved from https://ghr.nlm.nih.gov/.

Uemoto, Y., Nakano, H., Kikuchi, T., Sato, S., Ishida, M., Shibata, T., *et al.* (2012)

'Fine mapping of porcine SSC14 QTL and SCD gene effects on fatty acid composition and melting point of fat in a Duroc purebred population.', *Animal genetics*, 43(2), pp. 225–8.

VanRaden, P. M., O'Connell, J. R., Wiggans, G. R. and Weigel, K. A. (2011) 'Genomic evaluations with many more genotypes.', *Genetics, selection, evolution : GSE*, 43(1), p. 10.

Varel, V. H., Yen, J. T. and Kreikemeier, K. K. (1995) 'Addition of cellulolytic clostridia to the bovine rumen and pig intestinal tract.', *Applied and environmental microbiology*, 61(3), pp. 1116–9.

Varona, L., Ovilo, C., Clop, A., Noguera, J. L., Pérez-Enciso, M., Coll, A., *et al.* (2002) 'QTL mapping for growth and carcass traits in an Iberian by Landrace pig intercross: additive, dominant and epistatic effects.', *Genetical research*, 80(2), pp. 145–54.

Varona, L. (2017) 'Mejora genética porcina: más allá de la genómica', *Suis*, 137, pp. 12–17.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., *et al.* (2004) 'Environmental genome shotgun sequencing of the Sargasso Sea.', *Science (New York, N.Y.)*, 304(5667), pp. 66–74.

Viterbo, V. S., Lopez, B. I. M., Kang, H., Kim, H., Song, C. and Seo, K. S. (2018) 'Genome wide association study of fatty acid composition in Duroc swine.', *Asian-Australasian journal of animal sciences*, 31(8), pp. 1127–1133.

Wagner Mackenzie, B., Waite, D. W. and Taylor, M. W. (2015) 'Evaluating variation in human gut microbiota profiles due to DNA extraction method and inter-subject differences.', *Frontiers in microbiology*, 6(FEB), p. 130.

Wajner, M. and Amaral, A. U. (2015) 'Mitochondrial dysfunction in fatty acid oxidation disorders: insights from human and animal studies.', *Bioscience reports*, 36(1), p. e00281.

Wakil, S. J., Stoops, J. K. and Joshi, V. C. (1983) 'Fatty acid synthesis and its regulation.', *Annual review of biochemistry*, 52(1), pp. 537–79.

Wan, Y., Zhang, S., Li, L., Chen, H. and Zhou, R. (2017) 'Characterization of a

novel streptococcal heme-binding protein SntA and its interaction with host antioxidant protein AOP2.', *Microbial pathogenesis*, 111, pp. 145–155.

Wang, M. and Donovan, S. M. (2015) 'Human microbiota-associated swine: current progress and future opportunities.', *ILAR journal*, 56(1), pp. 63–73.

Ward, D. M., Weller, R. and Bateson, M. M. (1990) '16S rRNA sequences reveal numerous uncultured microorganisms in a natural community.', *Nature*, 345(6270), pp. 63–5.

Webb, E. C. and O'Neill, H. A. (2008) 'The animal fat paradox and meat quality.', *Meat science*, 80(1), pp. 28–36.

Weber, J. L., David, D., Heil, J., Fan, Y., Zhao, C. and Marth, G. (2002) 'Human Diallelic Insertion/Deletion Polymorphisms', *The American Journal of Human Genetics*, 71(4), pp. 854–862.

Wei, Z., Peterson, J. M., Lei, X., Cebotaru, L., Wolfgang, M. J., Baldeviano, G. C., *et al.* (2012) 'C1q/TNF-related Protein-12 (CTRP12), a Novel Adipokine That Improves Insulin Sensitivity and Glycemic Control in Mouse Models of Obesity and Diabetes', *Journal of Biological Chemistry*, 287(13), pp. 10301–10315.

Weisburg, W. G., Barns, S. M., Pelletier, D. A. and Lane, D. J. (1991) '16S ribosomal DNA amplification for phylogenetic study.', *Journal of bacteriology*, 173(2), pp. 697–703.

White, S. (2011) 'From Globalized Pig Breeds to Capitalist Pigs: A Study in Animal Cultures and Evolutionary History', *Environmental History*, 16(1), pp. 94–120.

Woese, C. R., Gutell, R., Gupta, R. and Noller, H. F. (1983) 'Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids.', *Microbiological reviews*, 47(4), pp. 621–69.

Woese, C. R. (1987) 'Bacterial evolution.', *Microbiological reviews*, 51(2), pp. 221–71.

Woese, C. R. and Fox, G. E. (1977) 'Phylogenetic structure of the prokaryotic domain: the primary kingdoms.', *Proceedings of the National Academy of*

Sciences of the United States of America, 74(11), pp. 5088–90.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Richardson, R. I. and Sheard, P. R. (1999) 'Manipulating meat quality and composition.', *The Proceedings of the Nutrition Society*, 58(2), pp. 363–70.

Wood, J. D., Richardson, R. I., Nute, G. R., Fisher, A. V, Campo, M. M., Kasapidou, E., *et al.* (2004) 'Effects of fatty acids on meat quality: a review.', *Meat science*, 66(1), pp. 21–32.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., *et al.* (2008) 'Fat deposition, fatty acid composition and meat quality: A review.', *Meat science*, 78(4), pp. 343–58.

Wood, J. and Whittemore, C. T. (2006) 'Pig meat and carcass quality', in Kyriazakis, I. and Whittemore, C. T. (eds) *Whittemore's Science and Practice of Pig Production*. 3rd edn. Oxford, UK: Blackwell Publishing Ltd.

Xiao, L., Estellé, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., *et al.* (2016) 'A reference gene catalogue of the pig gut microbiome', *Nature Microbiology*. Nature Publishing Group, 1(12), p. 16161.

Yan, H., Diao, H., Xiao, Y., Li, W., Yu, B., He, J., *et al.* (2016) 'Gut microbiota can transfer fiber characteristics and lipid metabolic profiles of skeletal muscle from pigs to germ-free mice.', *Scientific reports*. Nature Publishing Group, 6(1), p. 31786.

Yang, B., Zhang, W., Zhang, Z., Fan, Y., Xie, X., Ai, H., *et al.* (2013) 'Genome-wide association analyses for fatty acid composition in porcine muscle and abdominal fat tissues.', *PloS one*, 8(6), p. e65554.

Yang, H., Huang, X., Fang, S., Xin, W., Huang, L. and Chen, C. (2016) 'Uncovering the composition of microbial community structure and metagenomics among three gut locations in pigs with distinct fatness.', *Scientific reports*. Nature Publishing Group, 6(May), p. 27427.

Yang, J., Keshavarzian, A. and Rose, D. J. (2013) 'Impact of dietary fiber fermentation from cereal grains on metabolite production by the fecal microbiota from normal weight and obese individuals.', *Journal of medicinal food*, 16(9), pp. 862–7.

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., *et al.* (2014) 'Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences.', *Nature reviews. Microbiology*. Nature Publishing Group, 12(9), pp. 635–45.

Young, B., Lowe, J. S., Stevens, A. and Heath, J. W. (2006) *Wheater's Functional Histology: A Text and Colour Atlas*. 5th edn. Churchill Livingstone/Elsevier.

Zadravec, D., Brolinson, A., Fisher, R. M., Carneheim, C., Csikasz, R. I., Bertrand-Michel, J., *et al.* (2010) 'Ablation of the very-long-chain fatty acid elongase ELOVL3 in mice leads to constrained lipid storage and resistance to diet-induced obesity.', *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 24(11), pp. 4366–77.

Zang, L., Wang, Y., Sun, B., Zhang, X., Yang, C., Kang, L., *et al.* (2016) 'Identification of a 13 bp indel polymorphism in the 3'-UTR of DGAT2 gene associated with backfat thickness and lean percentage in pigs.', *Gene*, 576(2 Pt 2), pp. 729–33.

Zappaterra, M., Ros-Freixedes, R., Estany, J. and Davoli, R. (2018) 'Association study highlights the influence of ELOVL fatty acid elongase 6 gene region on backfat fatty acid composition in Large White pig breed.', *Animal*, 48(1), pp. 1–10.

Zhang, H., DiBaise, J. K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., *et al.* (2009) 'Human gut microbiota in obesity and after gastric bypass.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(7), pp. 2365–70.

Zhang, W., Zhang, J., Cui, L., Ma, J., Chen, C., Ai, H., *et al.* (2016a) 'Genetic architecture of fatty acid composition in the longissimus dorsi muscle revealed by genome-wide association studies on diverse pig populations.', *Genetics Selection Evolution*. BioMed Central, 48(1), p. 5.

Zhang, W., Bin Yang, Zhang, J., Cui, L., Ma, J., Chen, C., *et al.* (2016b) 'Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations.', *Scientific reports*. Nature Publishing Group, 6(1), p. 24718.

Zhao, W., Wang, Y., Liu, S., Huang, J., Zhai, Z., He, C., *et al.* (2015) 'The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments.', *PloS one*, 10(2), p. e0117441.

Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., *et al.* (2016) 'Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity.', *Science (New York, N.Y.)*, 352(6285), pp. 565–9.

Zoetendal, E. G., Raes, J., Bogert, B. Van Den, Booijink, C. C. G. M., Troost, F. J., Bork, P., *et al.* (2012) 'The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates.', *The ISME journal*. Nature Publishing Group, 6(7), pp. 1415–26.

7. ANNEXES

7.1 Supplementary material Paper I: 'Genome-Wide Association Study for backfat and intramuscular fatty acid composition in three different pig crosses based on the Iberian breed'

Supplementary Table S1. List of the significantly associated SNPs within QTL regions for the FA composition in backfat when GWAS were performed in the merged dataset and the predicted consequences of the SNPs.

Region	SNP	Chr	Position	MAF	*p*-value	FDR	Trait
BF1	rs80899816	1	146957212	0.172	1.38E-06	2.65E-02	C16:1(n-9)
BF1	rs80816307	1	146979672	0.18	9.22E-06	6.24E-02	C16:1(n-9)
BF2	rs81306755	2	145257	0.391	1.43E-05	6.24E-02	C16:1(n-9)
BF2	rs81306755	2	145257	0.391	2.35E-06	1.38E-02	C18:1(n-9)
BF2	rs81306755	2	145257	0.391	8.98E-09	2.97E-04	C18:2(n-6)
BF2	rs81306755	2	145257	0.391	1.35E-05	6.47E-02	MUFA
BF2	rs81306755	2	145257	0.391	3.07E-09	9.77E-05	MUFA/PUFA
BF2	rs81306755	2	145257	0.391	1.07E-08	3.65E-04	PUFA
BF2	rs81306755	2	145257	0.391	8.05E-07	1.82E-02	PUFA/SFA
BF2	rs81328276	2	236157	0.425	2.19E-06	1.38E-02	C18:1(n-9)
BF2	rs81328276	2	236157	0.425	8.69E-07	6.22E-03	C18:2(n-6)
BF2	rs81328276	2	236157	0.425	9.78E-06	6.27E-02	MUFA
BF2	rs81328276	2	236157	0.425	2.02E-07	1.11E-03	MUFA/PUFA
BF2	rs81328276	2	236157	0.425	1.12E-06	7.16E-03	PUFA
BF2	rs81317307	2	310819	0.442	4.40E-06	4.23E-02	C16:1(n-9)
BF2	rs81317307	2	310819	0.442	6.40E-06	2.23E-02	C18:1(n-9)
BF2	rs81317307	2	310819	0.442	1.55E-08	2.97E-04	C18:2(n-6)
BF2	rs81317307	2	310819	0.442	5.09E-09	9.77E-05	MUFA/PUFA
BF2	rs81317307	2	310819	0.442	1.90E-08	3.65E-04	PUFA
BF2	rs81317307	2	310819	0.442	9.47E-07	1.82E-02	PUFA/SFA
BF2	rs81339115	2	422634	0.451	3.57E-05	9.80E-02	C18:1(n-9)
BF2	rs81339115	2	422634	0.451	5.21E-06	1.82E-02	C18:2(n-6)
BF2	rs81339115	2	422634	0.451	2.82E-06	9.09E-03	MUFA/PUFA
BF2	rs81339115	2	422634	0.451	6.48E-06	2.26E-02	PUFA
BF2	rs81341763	2	677649	0.428	1.19E-06	1.38E-02	C18:1(n-9)
BF2	rs81341763	2	677649	0.428	2.79E-07	2.68E-03	C18:2(n-6)
BF2	rs81341763	2	677649	0.428	6.27E-06	5.13E-02	MUFA
BF2	rs81341763	2	677649	0.428	6.67E-08	5.12E-04	MUFA/PUFA
BF2	rs81341763	2	677649	0.428	3.67E-07	3.53E-03	PUFA
BF2	rs81341763	2	677649	0.428	1.76E-05	9.65E-02	PUFA/SFA
BF2	rs81291529	2	2636400	0.455	2.17E-05	6.73E-02	C16:1(n-9)

Region	SNP	Chr	Position	MAF	*p*-value	FDR	Trait
BF2	rs81356987	2	3859973	0.358	1.73E-05	6.66E-02	C16:1(n-9)
BF2	rs81356987	2	3859973	0.358	1.46E-06	8.01E-03	C18:2(n-6)
BF2	rs81356987	2	3859973	0.357	2.42E-06	9.09E-03	MUFA/PUFA
BF2	rs81356987	2	3859973	0.358	1.82E-06	8.75E-03	PUFA
BF2	rs81356987	2	3859973	0.358	9.09E-06	8.68E-02	PUFA/SFA
BF2	rs81357266	2	3985548	0.379	3.55E-06	1.51E-02	C18:1(n-9)
BF2	rs81357266	2	3985548	0.379	5.88E-06	1.88E-02	C18:2(n-6)
BF2	rs81357266	2	3985548	0.379	1.33E-06	5.68E-03	MUFA/PUFA
BF2	rs81357266	2	3985548	0.379	7.32E-06	2.35E-02	PUFA
BF2	rs81364067	2	4412670	0.448	2.74E-06	1.05E-02	C18:2(n-6)
BF2	rs81364067	2	4412670	0.448	4.73E-06	1.30E-02	MUFA/PUFA
BF2	rs81364067	2	4412670	0.448	2.67E-06	1.11E-02	PUFA
BF2	rs81364067	2	4412670	0.448	1.36E-05	8.68E-02	PUFA/SFA
BF2	rs81364734	2	4444553	0.451	2.38E-05	5.72E-02	C18:2(n-6)
BF2	rs81364734	2	4444553	0.451	1.22E-05	2.75E-02	MUFA/PUFA
BF2	rs81364734	2	4444553	0.451	2.39E-05	6.11E-02	PUFA
BF2	rs81368683	2	4966521	0.312	2.34E-05	6.73E-02	C16:1(n-9)
BF2	rs81368683	2	4966521	0.312	2.53E-07	2.68E-03	C18:2(n-6)
BF2	rs81368683	2	4966521	0.312	1.60E-07	1.02E-03	MUFA/PUFA
BF2	rs81368683	2	4966521	0.312	3.05E-07	3.53E-03	PUFA
BF2	rs81368683	2	4966521	0.312	3.29E-06	4.21E-02	PUFA/SFA
BF2	rs81346312	2	5372107	0.228	2.45E-05	6.73E-02	C16:1(n-9)
BF2	rs81356796	2	5827995	0.194	3.39E-05	9.80E-02	C18:1(n-9)
BF2	rs81356796	2	5827995	0.194	2.34E-05	5.72E-02	C18:2(n-6)
BF2	rs81356796	2	5827995	0.194	3.04E-05	6.87E-02	PUFA
BF2	rs81285769	2	5830742	0.166	1.60E-06	1.38E-02	C18:1(n-9)
BF2	rs81285769	2	5830742	0.166	1.84E-05	5.06E-02	C18:2(n-6)
BF2	rs81285769	2	5830742	0.165	1.98E-06	3.81E-02	MUFA
BF2	rs81285769	2	5830742	0.166	2.26E-05	6.11E-02	PUFA
BF2	rs81358530	2	6962060	0.392	2.48E-05	7.93E-02	C18:1(n-9)
BF2	rs81358530	2	6962060	0.392	2.55E-06	1.05E-02	C18:2(n-6)
BF2	rs81358530	2	6962060	0.393	1.25E-06	5.68E-03	MUFA/PUFA
BF2	rs81358530	2	6962060	0.392	2.90E-06	1.11E-02	PUFA
BF2	rs81359337	2	7852330	0.264	2.29E-06	1.38E-02	C18:1(n-9)
BF2	rs81359337	2	7852330	0.263	3.77E-06	4.83E-02	MUFA
BF2	rs81359616	2	8060705	0.412	4.93E-07	1.38E-02	C18:1(n-9)
BF2	rs81359616	2	8060705	0.412	1.15E-06	3.81E-02	MUFA
BF2	rs81359616	2	8060705	0.412	4.27E-06	1.26E-02	MUFA/PUFA
BF2	rs81360111	2	8647689	0.466	1.16E-05	6.24E-02	C16:1(n-9)

Region	SNP	Chr	Position	MAF	*p*-value	FDR	Trait
BF2	rs81360111	2	8647689	0.466	3.36E-06	1.51E-02	C18:1(n-9)
BF2	rs81360111	2	8647689	0.466	2.14E-06	1.03E-02	C18:2(n-6)
BF2	rs81360111	2	8647689	0.466	1.75E-05	7.48E-02	MUFA
BF2	rs81360111	2	8647689	0.466	3.72E-08	4.77E-04	MUFA/PUFA
BF2	rs81360111	2	8647689	0.466	1.74E-06	8.75E-03	PUFA
BF2	rs81474907	2	8795034	0.242	1.12E-05	2.69E-02	MUFA/PUFA
BF2	rs81312355	2	8894947	0.383	2.52E-06	1.38E-02	C18:1(n-9)
BF2	rs81312355	2	8894947	0.383	6.68E-06	5.13E-02	MUFA
BF2	rs81214179	2	8936150	0.362	5.84E-06	2.23E-02	C18:1(n-9)
BF2	rs81214179	2	8936150	0.362	9.72E-07	6.22E-03	C18:2(n-6)
BF2	rs81214179	2	8936150	0.362	1.16E-05	6.37E-02	MUFA
BF2	rs81214179	2	8936150	0.362	6.45E-08	5.12E-04	MUFA/PUFA
BF2	rs81214179	2	8936150	0.362	1.04E-06	7.16E-03	PUFA
BF2	rs81360570	2	9772447	0.259	4.79E-05	9.01E-02	MUFA/PUFA
BF3	rs45432640	4	81446371	0.299	3.06E-05	6.18E-02	MUFA/PUFA
BF3	rs80847745	4	81479518	0.44	5.16E-05	9.01E-02	MUFA/PUFA
BF3	rs80848071	4	81493481	0.415	6.01E-06	1.54E-02	MUFA/PUFA
BF3	rs81318065	4	81565259	0.334	2.04E-05	4.35E-02	MUFA/PUFA
BF4	rs81345002	6	16339713	0.376	1.75E-06	3.36E-02	C20:2(n-6)
BF4	rs81322046	6	16344565	0.203	3.46E-07	1.33E-02	C20:2(n-6)
BF4	rs81336454	6	26053724	0.139	6.69E-06	4.78E-02	C20:2(n-6)
BF5	rs81403348	8	109399818	0.159	1.28E-09	1.64E-05	C16:0
BF5	rs81403348	8	109399818	0.159	2.47E-05	5.58E-02	C14:0
BF5	rs81403349	8	109417642	0.211	3.05E-10	1.17E-05	C16:0
BF5	rs81403349	8	109417642	0.211	1.86E-05	8.92E-02	C16:1(n-7)
BF5	rs81403349	8	109417642	0.211	2.78E-05	5.63E-02	C14:0
BF5	rs81403355	8	109446670	0.215	6.25E-10	1.20E-05	C16:0
BF5	rs81403355	8	109446670	0.215	1.19E-05	3.05E-02	C14:0
BF5	rs81403355	8	109446670	0.215	1.62E-05	8.91E-02	C16:1(n-7)
BF5	rs81403368	8	109576850	0.196	2.34E-09	2.25E-05	C16:0
BF5	rs81403368	8	109576850	0.196	7.98E-07	1.29E-02	C16:1(n-7)
BF5	rs81403368	8	109576850	0.196	2.29E-06	7.33E-03	C14:0
BF5	rs339502807	8	111429584	0.319	8.84E-09	6.79E-05	C16:0
BF5	rs81403466	8	111899793	0.198	3.16E-08	2.02E-04	C16:0
BF5	rs81403466	8	111899793	0.198	9.89E-06	7.14E-02	C16:1(n-7)
BF5	rs81403466	8	111899793	0.198	1.58E-05	3.79E-02	C14:0
BF5	rs81310214	8	112005632	0.477	5.29E-06	5.08E-02	C16:1(n-7)
BF5	rs81293101	8	112877898	0.252	1.12E-05	7.14E-02	C16:1(n-7)
BF5	rs81477002	8	113809916	0.205	7.65E-06	3.67E-02	C16:0

Region	SNP	Chr	Position	MAF	*p*-value	FDR	Trait
BF5	rs81332214	8	114403588	0.382	7.88E-08	4.32E-04	C16:0
BF5	rs81332214	8	114403588	0.382	4.99E-05	9.59E-02	C14:0
BF5	rs81306425	8	114474492	0.394	1.57E-05	6.69E-02	C16:0
BF5	rs81292625	8	114483898	0.336	2.22E-05	8.52E-02	C16:0
BF5	rs81310709	8	115404032	0.422	1.01E-06	1.29E-02	C16:1(n-7)
BF5	rs81476888	8	115409757	0.468	4.95E-07	1.29E-02	C16:1(n-7)
BF5	rs81476888	8	115409757	0.468	1.00E-05	2.75E-02	C14:0
BF6	rs81266750	10	30495950	0.213	3.30E-05	7.92E-02	C16:1(n-9)
BF6	rs81423282	10	30742643	0.256	1.46E-05	6.24E-02	C16:1(n-9)
BF6	rs81423288	10	30756975	0.256	1.46E-05	6.24E-02	C16:1(n-9)
BF7	rs81289413	10	52874658	0.09	1.52E-05	7.45E-02	C20:2(n-6)
BF7	rs80979357	10	52920353	0.095	7.46E-06	4.78E-02	C20:2(n-6)
BF8	rs81271924	12	275081	0.183	5.04E-08	3.23E-04	C14:0
BF8	rs81440337	12	278538	0.182	4.58E-07	1.60E-03	C14:0
BF8	rs81312632	12	304547	0.169	1.63E-07	6.96E-04	C14:0
BF8	rs81327505	12	324004	0.169	1.27E-07	6.10E-04	C14:0
BF8	rs81434341	12	353286	0.168	1.11E-07	6.08E-04	C14:0
BF8	rs81439363	12	379031	0.168	2.41E-07	9.25E-04	C14:0
BF8	rs81433040	12	417230	0.201	2.87E-08	2.76E-04	C14:0
BF8	rs81266346	12	429636	0.204	1.05E-08	1.35E-04	C14:0
BF8	rs81271195	12	547357	0.219	3.22E-09	6.19E-05	C14:0
BF8	rs81311681	12	658326	0.247	3.86E-08	2.97E-04	C14:0
BF8	rs81308244	12	824959	0.237	1.98E-10	7.59E-06	C14:0
BF8	rs81286950	12	910198	0.115	5.36E-06	1.58E-02	C14:0
BF9	rs81457961	16	30669240	0.201	3.82E-05	5.87E-02	C20:1(n-9)/C20:0
BF9	rs81267800	16	38142493	0.094	5.63E-05	7.46E-02	C20:1(n-9)/C20:0
BF9	rs81297480	16	38349508	0.127	3.20E-08	6.14E-04	C20:1(n-9)/C20:0
BF9	rs81458871	16	38483905	0.127	3.20E-08	6.14E-04	C20:1(n-9)/C20:0
BF9	rs81458940	16	39357333	0.176	4.99E-06	1.60E-02	C20:1(n-9)/C20:0
BF9	rs81458993	16	40203541	0.156	1.12E-06	4.32E-03	C20:1(n-9)/C20:0
BF9	rs81459001	16	40402928	0.161	1.30E-06	4.53E-03	C20:1(n-9)/C20:0
BF9	rs81297382	16	40534440	0.147	4.40E-07	3.38E-03	C20:1(n-9)/C20:0
BF9	rs81321258	16	40995213	0.177	9.48E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs81243029	16	41483466	0.177	9.48E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs81297837	16	41757449	0.177	9.48E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs81459077	16	41900163	0.177	9.48E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs81459084	16	42103210	0.197	3.62E-05	5.87E-02	C20:1(n-9)/C20:0
BF9	rs81459081	16	42185492	0.177	9.48E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs81459087	16	42284127	0.177	9.48E-06	1.73E-02	C20:1(n-9)/C20:0

Region	SNP	Chr	Position	MAF	*p*-value	FDR	Trait
BF9	rs81297448	16	42409389	0.147	4.40E-07	3.38E-03	C20:1(n-9)/C20:0
BF9	rs81297627	16	42582565	0.147	4.40E-07	3.38E-03	C20:1(n-9)/C20:0
BF9	rs81459126	16	42763329	0.177	8.34E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs320616585	16	43727250	0.177	9.48E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs81297790	16	44809468	0.144	1.11E-06	4.32E-03	C20:1(n-9)/C20:0
BF9	rs81459297	16	44886367	0.227	9.31E-06	1.73E-02	C20:1(n-9)/C20:0
BF9	rs81459286	16	44908847	0.144	1.11E-06	4.32E-03	C20:1(n-9)/C20:0
BF9	rs81459319	16	45025316	0.144	1.11E-06	4.32E-03	C20:1(n-9)/C20:0
BF9	rs81247528	16	45119683	0.154	6.07E-05	7.78E-02	C20:1(n-9)/C20:0
BF9	rs81459345	16	45245630	0.144	1.11E-06	4.32E-03	C20:1(n-9)/C20:0
BF9	rs81305530	16	45586121	0.138	4.53E-05	6.22E-02	C20:1(n-9)/C20:0
BF9	rs81345265	16	45614639	0.138	4.53E-05	6.22E-02	C20:1(n-9)/C20:0
BF9	rs81277305	16	45621491	0.138	4.53E-05	6.22E-02	C20:1(n-9)/C20:0
BF9	rs81459434	16	45892779	0.142	3.76E-05	5.87E-02	C20:1(n-9)/C20:0
BF9	rs81459469	16	46175075	0.142	3.76E-05	5.87E-02	C20:1(n-9)/C20:0
BF9	rs81459502	16	46719890	0.135	7.73E-05	8.73E-02	C20:1(n-9)/C20:0
BF9	rs81459516	16	46938356	0.135	7.73E-05	8.73E-02	C20:1(n-9)/C20:0
BF9	rs81459530	16	47061035	0.135	7.73E-05	8.73E-02	C20:1(n-9)/C20:0
BF9	rs81224313	16	47628216	0.135	7.73E-05	8.73E-02	C20:1(n-9)/C20:0

Supplementary Table S1. continuation (*consequence prediction*)

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
BF1	rs80899816	intergenic_variant	-	-
BF1	rs80816307	intergenic_variant	-	-
BF2	rs81306755	intron_variant	ENSSSCG00000014565	-
BF2	rs81328276	intron_variant	ENSSSCG00000024569	ANO9
BF2	rs81317307	intron_variant	ENSSSCG00000027045	LRRC56
BF2	rs81339115	intron_variant	ENSSSCG00000012850	DEAF1
BF2	rs81341763	intergenic_variant	-	-
BF2	rs81291529	intron_variant	ENSSSCG00000033043	SHANK2
BF2	rs81356987	intergenic_variant	-	-
BF2	rs81357266	intergenic_variant	-	-
BF2	rs81364067	intron_variant	ENSSSCG00000012884	PPP6R3
BF2	rs81364734	intron_variant	ENSSSCG00000012884	PPP6R3
BF2	rs81368683	upstream_gene_variant	ENSSSCG00000012896	NDUFV1
BF2	rs81346312	intergenic_variant	-	-
BF2	rs81356796	intron_variant	ENSSSCG00000012934	CCS
BF2	rs81285769	upstream_gene_variant; intron_variant	ENSSSCG00000012933; ENSSSCG00000012934	CCDC87; CCS

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
BF2	rs81358530	intron_variant	ENSSSCG00000012999	CAPN1
BF2	rs81359337	intron_variant	ENSSSCG00000013034; ENSSSCG00000034755	PLCB3; -
BF2	rs81359616	intron_variant	ENSSSCG00000013043	MACROD1
BF2	rs81360111	intron_variant	ENSSSCG00000028537	-
BF2	rs81474907	intron_variant	ENSSSCG00000029516	SLC22A8
BF2	rs81312355	synonymous_variant	ENSSSCG00000022404	SLC3A2
BF2	rs81214179	upstream_gene_variant; 3_prime_UTR_variant, synonymous_variant; upstream_gene_variant	ENSSSCG00000026293; ENSSSCG00000028025; ENSSSCG00000022866	STX5; WDR74; NXF1
BF2	rs81360570	intron_variant	ENSSSCG00000013078	MYRF
BF3	rs45432640	intron_variant	ENSSSCG00000006289	F5
BF3	rs80847745	intron_variant	ENSSSCG00000006289	F5
BF3	rs80848071	intergenic_variant	-	-
BF3	rs81318065	intron_variant; upstream_gene_variant	ENSSSCG00000028804; ENSSSCG00000033915	CCDC181; -
BF4	rs81345002	intergenic_variant	-	-
BF4	rs81322046	intergenic_variant	-	-
BF4	rs81336454	intergenic_variant	-	-
BF5	rs81403348	intron_variant	ENSSSCG00000009125	ANK2
BF5	rs81403349	intron_variant	ENSSSCG00000009125	ANK2
BF5	rs81403355	intron_variant	ENSSSCG00000009125	ANK2
BF5	rs81403368	intergenic_variant	-	-
BF5	rs339502807	intergenic_variant	-	-
BF5	rs81403466	intergenic_variant	-	-
BF5	rs81310214	downstream_gene_variant; upstream_gene_variant	ENSSSCG00000009133; ENSSSCG00000032883	-; -
BF5	rs81293101	intron_variant	ENSSSCG00000039813	-
BF5	rs81477002	intergenic_variant	-	-
BF5	rs81332214	intergenic_variant	-	-
BF5	rs81306425	intergenic_variant	-	-
BF5	rs81292625	intergenic_variant	-	-
BF5	rs81310709	intergenic_variant	-	-
BF5	rs81476888	upstream_gene_variant	ENSSSCG00000039563	GIMD1
BF6	rs81266750	intergenic_variant	-	-
BF6	rs81423282	intergenic_variant	-	-
BF6	rs81423288	intergenic_variant	-	-
BF5	rs81306425	intergenic_variant	-	-
BF5	rs81292625	intergenic_variant	-	-
BF5	rs81310709	intergenic_variant	-	-
BF5	rs81476888	upstream_gene_variant	ENSSSCG00000039563	GIMD1

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
BF6	rs81266750	intergenic_variant	-	-
BF6	rs81423282	intergenic_variant	-	-
BF6	rs81423288	intergenic_variant	-	-
BF7	rs81289413	intron_variant	ENSSSCG00000026001	DNAJC1
BF7	rs80979357	intron_variant	ENSSSCG00000026001	DNAJC1
BF8	rs81271924	intron_variant	ENSSSCG00000027229	B3GNTL1
BF8	rs81440337	intron_variant	ENSSSCG00000027229	B3GNTL1
BF8	rs81312632	intron_variant	ENSSSCG00000027229	B3GNTL1
BF8	rs81327505	downstream_gene_variant; intron_variant	ENSSSCG00000017136; ENSSSCG00000027229	TBCD; B3GNTL1
BF8	rs81434341	intron_variant	ENSSSCG00000017136	TBCD
BF8	rs81439363	intron_variant	ENSSSCG00000017136	TBCD
BF8	rs81433040	intron_variant	ENSSSCG00000017136	TBCD
BF8	rs81266346	intron_variant	ENSSSCG00000017136	TBCD
BF8	rs81271195	3_prime_UTR_variant	ENSSSCG00000017133	WDR45B
BF8	rs81311681	intron_variant; downstream_gene_variant; intron_variant	ENSSSCG00000017127; ENSSSCG00000017128; ENSSSCG00000039807	CYBC1; -; -
BF8	rs81308244	intergenic_variant	-	-
BF8	rs81286950	intron_variant	ENSSSCG00000029553	-
BF9	rs81457961	intergenic_variant	-	-
BF9	rs81267800	intron_variant	ENSSSCG00000016929	PDE4D
BF9	rs81297480	intergenic_variant	-	-
BF9	rs81458871	intergenic_variant	-	-
BF9	rs81458940	intergenic_variant	-	-
BF9	rs81458993	intron_variant	ENSSSCG00000025194	ZSWIM6
BF9	rs81459001	intergenic_variant	-	-
BF9	rs81297382	intergenic_variant	-	-
BF9	rs81321258	intron_variant	ENSSSCG00000030186	KIF2A
BF9	rs81243029	intron_variant	ENSSSCG00000016938	IPO11
BF9	rs81297837	intergenic_variant	-	-
BF9	rs81459077	intergenic_variant	-	-
BF9	rs81459084	intergenic_variant	-	-
BF9	rs81459081	intergenic_variant	-	-
BF9	rs81459087	intergenic_variant	-	-
BF9	rs81297448	intergenic_variant	-	-
BF9	rs81297627	intergenic_variant	-	-
BF9	rs81459126	intron_variant	ENSSSCG00000016941	RNF180
BF9	rs320616585	intron_variant	ENSSSCG00000016943	ADAMTS6
BF9	rs81297790	intergenic_variant	-	-
BF9	rs81459297	intergenic_variant	-	-

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
BF9	rs81459286	intergenic_variant	-	-
BF9	rs81459319	intron_variant	ENSSSCG00000016956	MAST4
BF9	rs81247528	intron_variant	ENSSSCG00000016956	MAST4
BF9	rs81459345	intron_variant	ENSSSCG00000016956	MAST4
BF9	rs81305530	intron_variant	ENSSSCG00000016957	CD180
BF9	rs81345265	intergenic_variant	-	-
BF9	rs81277305	intergenic_variant	-	-
BF9	rs81459434	intergenic_variant	-	-
BF9	rs81459469	intergenic_variant	-	-
BF9	rs81459502	intergenic_variant	-	-
BF9	rs81459516	intergenic_variant	-	-
BF9	rs81459530	intergenic_variant	-	-
BF9	rs81224313	intron_variant; downstream_gene_variant	ENSSSCG00000021514; ENSSSCG00000027380	OCLN; RF00026

Supplementary Table S2. List of the significantly associated SNPs within QTL regions for the FA composition in IMF when GWAS were performed in the merged dataset and the predicted consequences of the SNPs.

Region	SNP	Chr	Position	MAF	*p*-value	FDR	Trait
LD1	rs80992684	4	20019566	0.193	7.49E-06	8.53E-02	C20:4(n-6)/C20:3(n-6)
LD1	rs80910044	4	20057452	0.177	2.36E-06	8.53E-02	C20:4(n-6)/C20:3(n-6)
LD2	rs80919464	4	123756546	0.189	1.20E-05	7.65E-02	C20:4(n-6)
LD2	rs81347340	4	123895116	0.437	1.18E-05	7.65E-02	C20:4(n-6)
LD2	rs80915252	4	123979309	0.437	1.18E-05	7.65E-02	C20:4(n-6)
LD3	rs81441414	13	176539436	0.295	1.07E-05	7.65E-02	C20:4(n-6)
LD3	rs81441592	13	180138105	0.066	1.81E-06	6.96E-02	C20:4(n-6)
LD3	rs81317617	13	180652057	0.071	7.25E-06	7.65E-02	C20:4(n-6)
LD4	rs336373909	14	110946218	0.089	3.28E-05	5.05E-02	C18:1(n-9)/C18:0
LD4	rs80996278	14	111061647	0.087	3.14E-05	5.05E-02	C18:1(n-9)/C18:0
LD4	rs81252243	14	111842298	0.109	1.33E-06	5.12E-03	C18:1(n-9)/C18:0
LD4	rs81252243	14	111842298	0.109	1.21E-05	3.58E-02	MUFA/SFA
LD4	rs81449647	14	111856909	0.09	3.21E-05	5.05E-02	C18:1(n-9)/C18:0
LD4	rs80835267	14	112184991	0.095	5.75E-06	2.01E-02	C18:1(n-9)/C18:0
LD4	rs80835267	14	112184991	0.095	1.55E-05	3.97E-02	MUFA/SFA
LD4	rs80791021	14	112329198	0.117	1.21E-07	7.79E-04	C18:1(n-9)/C18:0
LD4	rs80791021	14	112329198	0.117	1.10E-06	7.41E-03	MUFA/SFA
LD4	rs335655209	14	112536527	0.115	1.08E-07	7.79E-04	C18:1(n-9)/C18:0
LD4	rs335655209	14	112536527	0.115	1.14E-06	7.41E-03	MUFA/SFA
LD4	rs80815886	14	112683768	0.117	1.21E-07	7.79E-04	C18:1(n-9)/C18:0
LD4	rs80815886	14	112683768	0.117	1.10E-06	7.41E-03	MUFA/SFA
LD4	rs327079688	14	112737326	0.117	1.22E-07	7.79E-04	C18:1(n-9)/C18:0
LD4	rs327079688	14	112737326	0.117	1.16E-06	7.41E-03	MUFA/SFA
LD4	rs80800867	14	112815936	0.117	1.21E-07	7.79E-04	C18:1(n-9)/C18:0
LD4	rs80800867	14	112815936	0.117	1.10E-06	7.41E-03	MUFA/SFA
LD4	rs80809309	14	112900857	0.115	3.06E-07	1.31E-03	C18:1(n-9)/C18:0
LD4	rs80809309	14	112900857	0.115	1.95E-06	8.98E-03	MUFA/SFA
LD4	rs80946643	14	112968076	0.114	2.83E-07	1.31E-03	C18:1(n-9)/C18:0
LD4	rs80946643	14	112968076	0.114	2.10E-06	8.98E-03	MUFA/SFA
LD4	rs80940194	14	113009739	0.168	3.17E-05	5.05E-02	C18:1(n-9)/C18:0
LD4	rs80963224	14	113083223	0.113	2.47E-07	1.31E-03	C18:1(n-9)/C18:0
LD4	rs80963224	14	113083223	0.113	1.85E-06	8.98E-03	MUFA/SFA
LD4	rs80954420	14	113229750	0.117	1.21E-07	7.79E-04	C18:1(n-9)/C18:0
LD4	rs80954420	14	113229750	0.117	1.10E-06	7.41E-03	MUFA/SFA
LD4	rs80996084	14	113360161	0.152	1.50E-05	3.21E-02	C18:1(n-9)/C18:0
LD4	rs80887368	14	113380424	0.152	1.50E-05	3.21E-02	C18:1(n-9)/C18:0
LD4	rs80948585	14	113426095	0.153	1.65E-05	3.33E-02	C18:1(n-9)/C18:0

Region	SNP	Chr	Position	MAF	*p*-value	FDR	Trait
LD4	rs321989995	14	113435462	0.152	1.50E-05	3.21E-02	C18:1(n-9)/C18:0
LD4	rs80885716	14	113462893	0.152	1.50E-05	3.21E-02	C18:1(n-9)/C18:0
LD4	rs337005009	14	113621937	0.152	1.50E-05	3.21E-02	C18:1(n-9)/C18:0
LD5	rs318740977	14	141151934	0.083	3.56E-06	3.26E-02	C18:0
LD5	rs80814938	14	141173643	0.083	3.56E-06	3.26E-02	C18:0
LD5	rs80939048	14	141232780	0.081	8.82E-06	5.65E-02	C18:0
LD5	rs80883500	14	141309602	0.083	3.56E-06	3.26E-02	C18:0
LD5	rs80950819	14	141338056	0.082	4.24E-06	3.26E-02	C18:0
LD5	rs80837196	14	141374711	0.082	4.24E-06	3.26E-02	C18:0
LD6	rs81303625	17	31061857	0.08	7.35E-06	9.09E-02	C20:0
LD6	rs80938452	17	31478463	0.149	9.46E-06	9.09E-02	C20:0
LD6	rs324135473	17	31599336	0.148	2.63E-06	9.09E-02	C20:0
LD6	rs81268001	17	31867849	0.139	5.14E-06	9.09E-02	C20:0

Supplementary Table S2. continuation (*consequence prediction*)

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
LD1	rs80992684	intergenic_variant	-	-
LD1	rs80910044	intergenic_variant	-	-
LD2	rs80919464	intron_variant	ENSSSCG00000006894	FNBP1L
LD2	rs81347340	intergenic_variant	-	-
LD2	rs80915252	splice_region_variant, synonymous_variant	ENSSSCG00000006897	CCDC18
LD3	rs81441414	intergenic_variant	-	-
LD3	rs81441592	intergenic_variant	-	-
LD3	rs81317617	intron_variant	ENSSSCG00000024623	USP25
LD4	rs336373909	intergenic_variant	-	-
LD4	rs80996278	intron_variant	ENSSSCG00000010544	DNMBP
LD4	rs81252243	intron_variant	ENSSSCG00000010556	PAX2
LD4	rs81449647	intron_variant	ENSSSCG00000010556	PAX2
LD4	rs80835267	intergenic_variant	-	-
LD4	rs80791021	intergenic_variant	-	-
LD4	rs335655209	intron_variant	ENSSSCG00000010563	BTRC
LD4	rs80815886	intron_variant	ENSSSCG00000010566	FBXW4
LD4	rs327079688	upstream_gene_variant	ENSSSCG00000010566	FBXW4
LD4	rs80800867	downstream_gene_variant; upstream_gene_variant	ENSSSCG00000010568; ENSSSCG00000036146	NPM3; FGF8
LD4	rs80809309	intron_variant	ENSSSCG00000010571	ARMH3
LD4	rs80946643	intron_variant	ENSSSCG00000010571	ARMH3

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
LD4	rs80940194	intron_variant	ENSSSCG00000010571	ARMH3
LD4	rs80963224	splice_region_variant, non_coding_transcript_exon_variant	ENSSSCG00000010573	HPS6
LD4	rs80954420	3_prime_UTR_variant; downstream_gene_variant	ENSSSCG00000010577; ENSSSCG00000010578	ELOVL3; PITX3
LD4	rs80996084	intron_variant	ENSSSCG00000010579	GBF1
LD4	rs80887368	downstream_gene_variant	ENSSSCG00000010579	GBF1
LD4	rs80948585	intron_variant; upstream_gene_variant	ENSSSCG00000010581; ENSSSCG00000038504	PSD; FBXL15
LD4	rs321989995	upstream_gene_variant; intron_variant; downstream_gene_variant	ENSSSCG00000010581; ENSSSCG00000033359; ENSSSCG00000038504	PSD; CUEDC2; FBXL15
LD4	rs80885716	upstream_gene_variant; intron_variant	ENSSSCG00000010584; ENSSSCG00000033359	MFSD13A; CUEDC2
LD4	rs337005009	intron_variant	ENSSSCG00000010586	SUFU
LD5	rs318740977	intron_variant	ENSSSCG00000010770	KNDC1
LD5	rs80814938	intron_variant	ENSSSCG00000010770	KNDC1
LD5	rs80939048	downstream_gene_variant; upstream_gene_variant	ENSSSCG00000010773; ENSSSCG00000010772	TUBGCP2; ADAM8
LD5	rs80883500	intron_variant	ENSSSCG00000034701	CALY
LD5	rs80950819	upstream_gene_variant; 3_prime_UTR_variant	ENSSSCG00000010791; ENSSSCG00000038086	FUOM; ECHS1
LD5	rs80837196	intron_variant	ENSSSCG00000029088	-
LD6	rs81303625	intron_variant	ENSSSCG00000034913	-
LD6	rs80938452	intergenic_variant	-	-
LD6	rs324135473	intergenic_variant	-	-
LD6	rs81268001	intron_variant	ENSSSCG00000007143	MAVS

Supplementary Table S3. Summary of the GWAS results for the thirteen SNPs genotyped in the three backcrosses.

SNP name	Reference SNP ID number	Chr.	Position (bp)	MAF	*p*-value	FDR	Trait
ELOVL6:c.-533C>T	rs345025813	8	112038523	0.14	1.44E-07	5.53E-04	C14:0
ELOVL6:c.-533C>T	rs345025813	8	112038523	0.14	6.42E-12	1.23E-07	C16:0
ELOVL6:c.-533C>T	rs345025813	8	112038523	0.14	3.79E-07	5.45E-03	C16:1(n-7)
ELOVL6:c.-394G>A	rs322956047	8	112038663	0.13	7.06E-08	3.88E-04	C14:0
ELOVL6:c.-394G>A	rs322956047	8	112038663	0.13	2.43E-12	9.32E-08	C16:0
ELOVL6:c.-394G>A	rs322956047	8	112038663	0.13	2.60E-07	5.45E-03	C16:1(n-7)
ELOVL6:c.416C>T	-	8	112180004	0.21	2.16E-06	5.50E-03	C14:0
ELOVL6:c.416C>T	-	8	112180004	0.21	1.98E-10	2.54E-06	C16:0
ELOVL6:c.416C>T	-	8	112180004	0.21	2.46E-06	1.35E-02	C16:1(n-7)
ELOVL6:c.1922C>T	-	8	112186937	0.24	6.32E-07	1.73E-03	C14:0
ELOVL6:c.1922C>T	-	8	112186937	0.24	5.10E-09	2.45E-05	C16:0
ELOVL6:c.1922C>T	-	8	112186937	0.24	5.67E-07	5.45E-03	C16:1(n-7)
PLA2G12A:c.-380T>A	rs335191239	8	112483189	0.48	2.53E-05	4.43E-02	C14:0
PLA2G12A:c.-380T>A	rs335191239	8	112483189	0.48	3.27E-06	1.05E-02	C16:0
PLA2G12A:c.-380T>A	rs335191239	8	112483189	0.48	6.37E-06	2.72E-02	C16:1(n-7)
ELOVL7:c.*1432A>G	rs325490947	16	39589566	0.18	1.80E-07	2.31E-03	C20:1(n-9)/C20:0
PIK3R1:c.-1256A>G	rs322671019	16	46434138	0.16	2.09E-05	3.49E-02	C20:1(n-9)/C20:0
PIK3R1:c.472A>G	rs331708297	16	46496141	0.15	5.76E-05	6.91E-02	C20:1(n-9)/C20:0

Supplementary Table S4. Frequencies of the thirteen SNPs genotyped in the three backcrosses.

SNP name	Reference SNP ID number	Chr.	Position (bp)	A1	A2	Animals	MAF	Freq (HOM A1)	Freq (HET)	Freq (HOM A2)
ELOVL6:c.-574T>C	rs325491325	8	112038483	C	T	3BCs	0.42	0.14	0.56	0.3
ELOVL6:c.-574T>C	rs325491325	8	112038483	C	T	BC1_DU	0.2	0.01	0.36	0.62
ELOVL6:c.-574T>C	rs325491325	8	112038483	C	T	BC1_LD	0.48	0.19	0.58	0.23
ELOVL6:c.-574T>C	rs325491325	8	112038483	C	T	BC1_PI	0.41	0.22	0.74	0.04
ELOVL6:c.-533C>T	rs345025813	8	112038523	T	C	3BCs	0.14	0.04	0.18	0.77
ELOVL6:c.-533C>T	rs345025813	8	112038523	T	C	BC1_DU	0	0	0	1
ELOVL6:c.-533C>T	rs345025813	8	112038523	T	C	BC1_LD	0.37	0.12	0.5	0.38
ELOVL6:c.-533C>T	rs345025813	8	112038523	T	C	BC1_PI	0	0	0	1
ELOVL6:c.-480C>T	rs341847499	8	112038577	T	C	3BCs	0.37	0.25	0.24	0.51
ELOVL6:c.-480C>T	rs341847499	8	112038577	T	C	BC1_DU	0.33	0.27	0.12	0.61
ELOVL6:c.-480C>T	rs341847499	8	112038577	T	C	BC1_LD	0.48	0.21	0.54	0.25
ELOVL6:c.-480C>T	rs341847499	8	112038577	T	C	BC1_PI	0.28	0.28	0	0.72
ELOVL6:c.-394G>A	rs322956047	8	112038663	A	G	3BCs	0.13	0.04	0.18	0.78
ELOVL6:c.-394G>A	rs322956047	8	112038663	A	G	BC1_DU	0	0	0	1
ELOVL6:c.-394G>A	rs322956047	8	112038663	A	G	BC1_LD	0.37	0.12	0.49	0.39
ELOVL6:c.-394G>A	rs322956047	8	112038663	A	G	BC1_PI	0	0	0	1
ELOVL6:c.416C>T	---	8	112180004	C	T	3BCs	0.21	0.08	0.27	0.65
ELOVL6:c.416C>T	---	8	112180004	C	T	BC1_DU	0.2	0.04	0.31	0.65
ELOVL6:c.416C>T	---	8	112180004	C	T	BC1_LD	0.41	0.17	0.48	0.36
ELOVL6:c.416C>T	---	8	112180004	C	T	BC1_PI	0	0	0	1
ELOVL6:c.1408C>T	---	8	112186423	G	A	3BCs	0.25	0.03	0.44	0.53
ELOVL6:c.1408C>T	---	8	112186423	G	A	BC1_DU	0.33	0.07	0.53	0.4
ELOVL6:c.1408C>T	---	8	112186423	G	A	BC1_LD	0.28	0.03	0.51	0.47
ELOVL6:c.1408C>T	---	8	112186423	G	A	BC1_PI	0.14	0	0.28	0.72
ELOVL6:c.1922C>T	---	8	112186937	A	G	3BCs	0.24	0.09	0.29	0.62
ELOVL6:c.1922C>T	---	8	112186937	A	G	BC1_DU	0.22	0.06	0.31	0.63
ELOVL6:c.1922C>T	---	8	112186937	A	G	BC1_LD	0.46	0.19	0.53	0.28
ELOVL6:c.1922C>T	---	8	112186937	A	G	BC1_PI	0	0	0	1
PLA2G12A:c.-380T>A	rs335191239	8	112483189	T	A	3BCs	0.48	0.24	0.49	0.27
PLA2G12A:c.-380T>A	rs335191239	8	112483189	T	A	BC1_DU	0.48	0.16	0.64	0.21
PLA2G12A:c.-380T>A	rs335191239	8	112483189	T	A	BC1_LD	0.41	0.2	0.41	0.39
PLA2G12A:c.-380T>A	rs335191239	8	112483189	T	A	BC1_PI	0.42	0.36	0.43	0.21
ELOVL7:c.*1432A>G	rs325490947	16	39589566	G	A	3BCs	0.18	0.05	0.26	0.69
ELOVL7:c.*1432A>G	rs325490947	16	39589566	G	A	BC1_DU	0	0	0	1
ELOVL7:c.*1432A>G	rs325490947	16	39589566	G	A	BC1_LD	0.39	0.13	0.53	0.34
ELOVL7:c.*1432A>G	rs325490947	16	39589566	G	A	BC1_PI	0.11	0	0.23	0.77

SNP name	Reference SNP ID number	Chr.	Position (bp)	A1	A2	Animals	MAF	Freq (HOM A1)	Freq (HET)	Freq (HOM A2)
ELOVL7:c.-46A>G	rs343494956	16	39680072	A	G	3BCs	0.4	0.13	0.54	0.33
ELOVL7:c.-46A>G	rs343494956	16	39680072	A	G	BC1_DU	0.46	0.13	0.65	0.22
ELOVL7:c.-46A>G	rs343494956	16	39680072	A	G	BC1_LD	0.47	0.21	0.65	0.14
ELOVL7:c.-46A>G	rs343494956	16	39680072	A	G	BC1_PI	0.18	0.02	0.31	0.67
ELOVL7:c.-629A>G	rs322657523	16	39680655	G	A	3BCs	0.27	0.05	0.45	0.5
ELOVL7:c.-629A>G	rs322657523	16	39680655	G	A	BC1_DU	0.29	0	0.57	0.43
ELOVL7:c.-629A>G	rs322657523	16	39680655	G	A	BC1_LD	0.44	0.13	0.63	0.25
ELOVL7:c.-629A>G	rs322657523	16	39680655	G	A	BC1_PI	0.06	0	0.12	0.88
PIK3R1:c.-1256A>G	rs322671019	16	46434138	C	T	3BCs	0.16	0.05	0.21	0.74
PIK3R1:c.-1256A>G	rs322671019	16	46434138	C	T	BC1_DU	0	0	0	1
PIK3R1:c.-1256A>G	rs322671019	16	46434138	C	T	BC1_LD	0.38	0.14	0.48	0.38
PIK3R1:c.-1256A>G	rs322671019	16	46434138	C	T	BC1_PI	0.05	0	0.1	0.9
PIK3R1:c.472A>G	rs331708297	16	46496141	C	T	3BCs	0.15	0.05	0.21	0.74
PIK3R1:c.472A>G	rs331708297	16	46496141	C	T	BC1_DU	0	0	0	1
PIK3R1:c.472A>G	rs331708297	16	46496141	C	T	BC1_LD	0.38	0.14	0.48	0.38
PIK3R1:c.472A>G	rs331708297	16	46496141	C	T	BC1_PI	0.05	0	0.11	0.89

Supplementary Table S5. Summary of the associated regions found in the three backcrosses when GWAS were performed individually.

Region	Backcross	Trait	Chr	Start	End	Trait
BC1_DU_BF1	BC1_DU	BF	1	229,067,988	242,017,170	C171; C200
BC1_DU_BF2	BC1_DU	BF	1	251,882,019	257,980,274	C200
BC1_DU_BF3	BC1_DU	BF	4	8,098,415	16,131,084	C204n6
BC1_DU_BF4	BC1_DU	BF	6	25,053,724	27,957,431	SFA
BC1_DU_BF5	BC1_DU	BF	12	0	7,855,894	C140
BC1_DU_LD1	BC1_DU	LD	1	116,193,074	118,242,644	C181n7/C161n7
BC1_DU_LD2	BC1_DU	LD	1	175,616,330	188,852,669	C201n9/C200
BC1_DU_LD3	BC1_DU	LD	2	19,157,764	24,225,758	C170
BC1_DU_LD4	BC1_DU	LD	2	36,783,291	47,457,134	C140
BC1_DU_LD5	BC1_DU	LD	2	86,511,858	92,459,562	C203n6; C204n6; PUFA; PUFA/SFA
BC1_DU_LD6	BC1_DU	LD	2	103,087,022	110,496,177	C204n6; PUFA/SFA
BC1_DU_LD7	BC1_DU	LD	3	121,844,545	125,004,026	PUFA/SFA
BC1_DU_LD8	BC1_DU	LD	4	13,919,602	16,131,084	C181n7/C161n7
BC1_DU_LD9	BC1_DU	LD	4	63,774,763	95,600,438	C161n9; C170; C183n3; C202n6
BC1_DU_LD10	BC1_DU	LD	5	35,314,172	37,346,640	C170
BC1_DU_LD11	BC1_DU	LD	6	27,394,342	29,988,781	C170
BC1_DU_LD12	BC1_DU	LD	6	39,301,937	43,998,265	C204n6; PUFA/SFA
BC1_DU_LD13	BC1_DU	LD	6	54,179,566	59,511,405	C204n6; PUFA/SFA
BC1_DU_LD14	BC1_DU	LD	6	93,360,155	98,104,675	C204n6; PUFA/SFA
BC1_DU_LD15	BC1_DU	LD	9	124,173,834	127,061,970	C202n6
BC1_DU_LD16	BC1_DU	LD	11	48,735,789	52,306,112	C161n9
BC1_DU_LD17	BC1_DU	LD	11	61,642,748	76,842,119	C161n9; C170; C171; C202n6
BC1_DU_LD18	BC1_DU	LD	13	174,351,419	177,512,257	C170
BC1_DU_LD19	BC1_DU	LD	13	186,116,419	188,385,735	C161n9; C170; C171; C201n9/C200; C203n3
BC1_DU_LD20	BC1_DU	LD	14	47,523,085	56,839,455	C170
BC1_DU_LD21	BC1_DU	LD	14	68,620,139	73,254,375	C161n9; C170; C202n6
BC1_DU_LD22	BC1_DU	LD	14	86,206,824	89,050,196	C170; C171
BC1_DU_LD23	BC1_DU	LD	14	108,404,872	115,622,115	C181n9/C180; MUFA/SFA
BC1_LD_BF1	BC1_LD	BF	2	2,062,975	10,548,890	C161n9; C182n6; MUFA/PUFA; PUFA; PUFA/SFA
BC1_LD_BF2	BC1_LD	BF	2	89,362,907	102,211,529	C181n7; C182n6; MUFA/PUFA; PUFA
BC1_LD_BF3	BC1_LD	BF	2	147,798,621	149,806,518	C183n3
BC1_LD_BF4	BC1_LD	BF	4	34,029,394	42,264,727	C161n9
BC1_LD_BF5	BC1_LD	BF	4	52,618,844	119,595,949	C161n9; MUFA/PUFA
BC1_LD_BF6	BC1_LD	BF	6	12,143,758	14,908,922	C182n6/C183n3; C200
BC1_LD_BF7	BC1_LD	BF	6	164,590,160	166,657,661	C161n9

Region	Backcross	Trait	Chr	Start	End	Trait
BC1_LD_BF8	BC1_LD	BF	7	43,524,117	49,168,881	C182n6/C183n3
BC1_LD_BF9	BC1_LD	BF	8	69,980,587	94,376,804	C160; C161n7
BC1_LD_BF10	BC1_LD	BF	8	104,171,530	123,210,438	C140; C160; C161n7; C181n7/C161n7
BC1_LD_BF11	BC1_LD	BF	11	7,940,153	10,103,118	C201n9/C200
BC1_LD_BF12	BC1_LD	BF	11	49,354,601	51,390,969	C182n6/C183n3
BC1_LD_BF13	BC1_LD	BF	12	47,181,577	49,617,457	C161n9
BC1_LD_BF14	BC1_LD	BF	14	22,403,217	25,145,966	C201n9/C200
BC1_LD_BF15	BC1_LD	BF	14	38,955,238	52,931,273	C201n9/C200
BC1_LD_BF16	BC1_LD	BF	16	33,817,847	56,296,400	C201n9/C200
BC1_LD_BF17	BC1_LD	BF	17	44,617,105	46,715,749	C161n9
BC1_LD_BF18	BC1_LD	BF	18	24,935,702	27,485,940	C182n6/C183n3
BC1_LD_BF19	BC1_LD	BF	X	39,586,473	42,391,576	C181n9; MUFA
BC1_LD_LD1	BC1_LD	LD	1	139,051,084	141,296,640	C204n6/C203n6
BC1_LD_LD2	BC1_LD	LD	2	4,827,995	6,830,742	C204n6/C203n6
BC1_LD_LD3	BC1_LD	LD	2	87,052,459	101,249,581	C203n3
BC1_LD_LD4	BC1_LD	LD	2	144,019,988	148,968,629	C181n9
BC1_LD_LD5	BC1_LD	LD	3	47,941,152	50,294,422	MUFA
BC1_LD_LD6	BC1_LD	LD	3	102,453,876	106,873,336	C204n6/C203n6
BC1_LD_LD7	BC1_LD	LD	4	1,672,184	10,451,581	C204n6/C203n6
BC1_LD_LD8	BC1_LD	LD	4	18,982,312	45,319,537	C181n9; C182n6; MUFA; MUFA/PUFA
BC1_LD_LD9	BC1_LD	LD	4	54,281,306	69,417,452	C181n9; C182n6; MUFA; MUFA/SFA
BC1_LD_LD10	BC1_LD	LD	4	104,460,669	107,074,548	C204n6/C203n6
BC1_LD_LD11	BC1_LD	LD	4	120,475,774	129,112,246	C161n7
BC1_LD_LD12	BC1_LD	LD	6	0	4,630,324	C181n9; MUFA/SFA
BC1_LD_LD13	BC1_LD	LD	6	11,625,397	19,215,210	C204n6/C203n6
BC1_LD_LD14	BC1_LD	LD	6	78,872,328	83,530,240	C181n9; MUFA; MUFA/SFA
BC1_LD_LD15	BC1_LD	LD	6	127,922,390	132,459,409	C204n6/C203n6
BC1_LD_LD16	BC1_LD	LD	7	89,122,140	93,399,480	C201n9/C200
BC1_LD_LD17	BC1_LD	LD	7	101,584,950	105,701,201	C204n6/C203n6
BC1_LD_LD18	BC1_LD	LD	8	86,268,315	94,376,804	C161n7
BC1_LD_LD19	BC1_LD	LD	8	105,344,539	114,809,916	C160; C161n7; C181n7/C161n7
BC1_LD_LD20	BC1_LD	LD	8	122,666,402	124,703,884	C203n3
BC1_LD_LD21	BC1_LD	LD	9	5,389,263	14,203,773	C204n6/C203n6
BC1_LD_LD22	BC1_LD	LD	9	18,694,989	22,055,971	C201n9/C200
BC1_LD_LD23	BC1_LD	LD	9	65,947,861	68,527,807	C203n3
BC1_LD_LD24	BC1_LD	LD	9	92,696,383	103,459,200	C161n9
BC1_LD_LD25	BC1_LD	LD	10	6,725,337	8,957,412	C171; C200
BC1_LD_LD26	BC1_LD	LD	11	4,591,824	11,387,046	C181n9/C180; C181n9; MUFA; MUFA/SFA

Region	Backcross	Trait	Chr	Start	End	Trait
BC1_LD_LD27	BC1_LD	LD	11	46,323,113	58,273,288	C181n9/C180; C204n6/C203n6; MUFA; MUFA/SFA
BC1_LD_LD28	BC1_LD	LD	17	12,257,304	15,909,080	C201n9/C200
BC1_LD_LD29	BC1_LD	LD	17	17,319,097	22,682,144	C171
BC1_LD_LD30	BC1_LD	LD	17	28,925,790	32,599,336	C200
BC1_LD_LD31	BC1_LD	LD	17	55,934,629	58,137,723	C203n3
BC1_LD_LD32	BC1_LD	LD	18	44,357,970	51,874,905	C182n6/C183n3
BC1_LD_LD33	BC1_LD	LD	X	119,692,562	124,006,430	MUFA
BC1_PI_BF1	BC1_PI	BF	1	145,804,205	148,001,436	C201n9/C200
BC1_PI_BF2	BC1_PI	BF	1	170,742,785	173,988,615	C160; C161n9; C180; C181n9/C180; C182n6; C183n3; MUFA/SFA; PUFA; PUFA/SFA; SFA
BC1_PI_BF3	BC1_PI	BF	2	0	10,772,447	C170; C181n9; C182n6; MUFA; MUFA/PUFA; PUFA
BC1_PI_BF4	BC1_PI	BF	6	13,412,764	18,785,071	C160
BC1_PI_BF5	BC1_PI	BF	6	25,053,724	27,074,710	C140
BC1_PI_BF6	BC1_PI	BF	6	46,685,552	48,715,812	C182n6; MUFA/PUFA; PUFA; PUFA/SFA
BC1_PI_BF7	BC1_PI	BF	10	48,786,714	52,235,724	C200
BC1_PI_LD1	BC1_PI	LD	1	45,476,758	55,614,559	C204n6; MUFA/SFA; PUFA/SFA
BC1_PI_LD2	BC1_PI	LD	1	74,630,916	76,772,982	C204n6; PUFA/SFA
BC1_PI_LD3	BC1_PI	LD	2	85,556,317	96,595,645	MUFA/SFA
BC1_PI_LD4	BC1_PI	LD	2	136,670,952	149,914,582	C181n9; MUFA; MUFA/SFA
BC1_PI_LD5	BC1_PI	LD	3	18,351,190	21,010,051	C160
BC1_PI_LD6	BC1_PI	LD	3	111,483,140	114,537,308	C183n3
BC1_PI_LD7	BC1_PI	LD	4	10,498,072	21,057,452	C181n9; C182n6; C202n6/C182n6; MUFA; MUFA/SFA; PUFA; PUFA/SFA
BC1_PI_LD8	BC1_PI	LD	5	8,491,168	10,503,106	C181n9/C180; C182n6/C183n3; MUFA/SFA
BC1_PI_LD9	BC1_PI	LD	5	29,783,586	32,063,327	C160
BC1_PI_LD10	BC1_PI	LD	5	49,272,210	58,064,470	C181n9; C204n6/C182n6; C204n6; MUFA; PUFA; PUFA/SFA
BC1_PI_LD11	BC1_PI	LD	6	19,724,134	22,493,885	C170
BC1_PI_LD12	BC1_PI	LD	6	99,547,077	101,551,392	C182n6/C183n3; C202n6/C182n6; C204n6; PUFA/SFA
BC1_PI_LD13	BC1_PI	LD	6	110,476,238	120,223,012	MUFA/SFA
BC1_PI_LD14	BC1_PI	LD	7	41,952,142	45,424,101	PUFA/SFA
BC1_PI_LD15	BC1_PI	LD	8	6,236,422	8,288,655	C204n6/C203n6
BC1_PI_LD16	BC1_PI	LD	9	5,996,326	8,592,969	C170
BC1_PI_LD17	BC1_PI	LD	9	125,777,906	130,172,751	C181n9/C180; C182n6/C183n3; C202n6/C182n6; MUFA/SFA
BC1_PI_LD18	BC1_PI	LD	9	130,775,259	132,917,032	C140

Region	Backcross	Trait	Chr	Start	End	Trait
BC1_PI_LD19	BC1_PI	LD	10	23,879,474	28,918,154	C202n6/C182n6; MUFA/SFA
BC1_PI_LD20	BC1_PI	LD	11	65,558,117	71,874,350	C160
BC1_PI_LD21	BC1_PI	LD	12	48,725,299	56,447,604	C140; C183n3
BC1_PI_LD22	BC1_PI	LD	13	42,600,291	48,297,528	C160
BC1_PI_LD23	BC1_PI	LD	13	68,618,284	126,323,265	C160
BC1_PI_LD24	BC1_PI	LD	13	175,169,597	191,267,573	C160
BC1_PI_LD25	BC1_PI	LD	14	5,275,594	8,815,388	C204n6/C203n6
BC1_PI_LD26	BC1_PI	LD	14	9,743,899	13,822,777	C202n6/C182n6
BC1_PI_LD27	BC1_PI	LD	14	18,550,452	24,614,640	MUFA/SFA
BC1_PI_LD28	BC1_PI	LD	14	36,414,113	38,452,275	MUFA/PUFA
BC1_PI_LD29	BC1_PI	LD	15	29,197,850	31,413,406	C140
BC1_PI_LD30	BC1_PI	LD	15	122,171,616	124,311,817	C140
BC1_PI_LD31	BC1_PI	LD	15	125,971,901	129,560,678	C204n6/C203n6; C204n6; MUFA; PUFA; PUFA/SFA
BC1_PI_LD32	BC1_PI	LD	16	18,583,421	20,669,219	C160
BC1_PI_LD33	BC1_PI	LD	16	50,530,344	62,203,701	C203n6/C182n6; MUFA/SFA
BC1_PI_LD34	BC1_PI	LD	16	70,302,000	77,250,807	C204n6/C203n6; MUFA/SFA
BC1_PI_LD35	BC1_PI	LD	17	44,223,215	56,550,577	C140; C182n6/C183n3; C202n6/C182n6; C204n6/C203n6; C204n6; MUFA/SFA; PUFA; PUFA/SFA
BC1_PI_LD36	BC1_PI	LD	X	95,383,417	97,704,995	C140

7.2 Supplementary material Paper II: 'Indel detection from Whole Genome Sequencing data and association with lipid metabolism in pigs'

Table S1. GEMMA output for the suggestive (FDR≤0.1) SNPs found in the GWAS analysis for the log2 normalization of the relative abundance of eicosadienoic acid in the *Longissimus dorsi* muscle of the BC1_PI population.

Polymorphism	Chr.	Position	MAF	*p*-value	FDR
rs81230425	3	68889478	0.142	5.77E-05	9.07E-02
rs80800030	5	84722868	0.051	4.96E-05	8.52E-02
rs81323853	6	36102300	0.416	4.78E-06	5.34E-02
rs81233173	6	39859759	0.234	1.23E-05	5.34E-02
rs81337470	6	41646557	0.445	8.07E-06	5.34E-02
rs80859145	6	45854092	0.215	3.02E-05	6.71E-02
rs333501942	6	46858013	0.38	3.83E-05	7.28E-02
rs329085840	6	50412233	0.223	1.70E-05	5.34E-02
rs81285488	6	53880159	0.427	2.15E-05	5.97E-02
C1QTNF12	6	63549854	0.221	1.77E-05	5.34E-02
rs81332544	6	64432983	0.226	3.34E-05	6.71E-02
rs81335919	6	65514681	0.383	1.63E-05	5.34E-02
rs81326136	6	65721779	0.383	1.63E-05	5.34E-02
rs81341284	6	65919016	0.383	1.63E-05	5.34E-02
rs81335441	6	66009991	0.383	1.63E-05	5.34E-02
rs81265819	6	66083353	0.383	1.63E-05	5.34E-02
rs81330392	6	66259117	0.23	4.83E-05	8.52E-02
rs81476483	6	66475413	0.347	2.60E-05	6.71E-02
rs81270239	6	66829184	0.383	1.63E-05	5.34E-02
rs81270030	6	70428427	0.354	9.55E-06	5.34E-02
rs80902484	6	73996865	0.391	3.17E-05	6.71E-02
rs81328033	6	74762302	0.347	2.85E-05	6.71E-02
rs80864602	15	44497429	0.113	5.18E-05	8.52E-02

7.3 Supplementary material Paper III: 'Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions'

Supplementary Table S1. This table contains the results of the presence/absence analysis at the genus level performed with metagenomeSeq[24]. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which genus is uniquely present in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon.

Duodenum vs jejunum

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Actinomycetaceae; g__	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Brevibacteriaceae; g__Brevibacterium	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Dermabacteraceae; g__Brachybacterium	Inf	5.2	Inf	1.08E-04	5.47E-04
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Microbacteriaceae; g__Leucobacter	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Micrococcaceae; g__Arthrobacter	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Yaniellaceae; g__Yaniella	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Porphyromonadaceae; g__	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Porphyromonadaceae; g__Porphyromonas	Inf	2.85	Inf	1.65E-03	5.23E-03
	k__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__	Inf	3.82	Inf	4.58E-04	1.82E-03
	k__Bacteria; p__Cyanobacteria; c__4C0d-2; o__YS2; f__; g__	Inf	3.82	Inf	4.58E-04	1.82E-03
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae; g__	Inf	2.85	Inf	1.65E-03	5.23E-03

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae; g__Natronobacillus	Inf	3.82	Inf	4.58E-04	1.82E-03
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Planococcaceae; g__	Inf	5.2	Inf	1.08E-04	5.47E-04
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Planococcaceae; g__Rummeliibacillus	Inf	3.82	Inf	4.58E-04	1.82E-03
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Planococcaceae; g__Sporosarcina	Inf	3.82	Inf	4.58E-04	1.82E-03
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Staphylococcaceae; g__Jeotgalicoccus	Inf	19.68	Inf	1.92E-07	4.07E-06
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Gemellales; f__Gemellaceae; g__Gemella	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Aerococcaceae; g__	Inf	19.68	Inf	1.92E-07	4.07E-06
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Enterococcaceae; g__	Inf	2.85	Inf	1.65E-03	5.23E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__Anaerococcus	Inf	19.68	Inf	1.92E-07	4.07E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__GW-34	Inf	11.01	Inf	2.69E-06	3.42E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__ph2	Inf	2.85	Inf	1.65E-03	5.23E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__Tissierella_Soehngenia	Inf	2.85	Inf	1.65E-03	5.23E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Proteiniclasticum	Inf	2.85	Inf	1.65E-03	5.23E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__	Inf	11.01	Inf	2.69E-06	3.42E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Coprococcus	Inf	3.82	Inf	4.58E-04	1.82E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Roseburia	Inf	19.68	Inf	1.92E-07	4.07E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Shuttleworthia	Inf	2.85	Inf	1.65E-03	5.23E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__	Inf	19.68	Inf	1.92E-07	4.07E-06

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Faecalibacterium	Inf	19.68	Inf	1.92E-07	4.07E-06
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Bulleidia	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Erysipelothrix	Inf	3.82	Inf	4.58E-04	1.82E-03
	k__Bacteria; p__Fusobacteria; c__Fusobacteriia; o__Fusobacteriales; f__Fusobacteriaceae; g__Fusobacterium	Inf	7.31	Inf	2.02E-05	1.28E-04
	k__Bacteria; p__Planctomycetes; c__Planctomycetia; o__Pirellulales; f__Pirellulaceae; g__	Inf	2.85	Inf	1.65E-03	5.23E-03
	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__Agrobacterium	Inf	5.2	Inf	1.08E-04	5.47E-04
	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhodobacterales; f__Rhodobacteraceae; g__	Inf	5.2	Inf	1.08E-04	5.47E-04
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter	Inf	11.01	Inf	2.69E-06	3.42E-05
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Actinomycetaceae; g__Actinomyces	0	0	0.09	2.69E-06	3.42E-05
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Catenibacterium	0	0	0.19	1.08E-04	5.47E-04

Jenunum vs ileum

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Actinomycetaceae; g__Actinomyces	Inf	11.01	Inf	2.69E-06	8.61E-05
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__[Paraprevotellaceae]; g__[Prevotella]	Inf	11.01	Inf	2.69E-06	8.61E-05
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__	Inf	3.82	Inf	4.58E-04	4.88E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia	Inf	7.31	Inf	2.02E-05	4.31E-04
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Catenibacterium	Inf	5.2	Inf	1.08E-04	1.38E-03
	k__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Flexispira	0	0	0.19	1.08E-04	1.38E-03

Ileum vs proximal colon

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Dietziaceae; g__Dietzia	Inf	11.01	Inf	2.69E-06	7.66E-06
	k__Bacteria; p__Cyanobacteria; c__Chloroplast; o__Streptophyta; f__; g__	Inf	11.01	Inf	2.69E-06	7.66E-06
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Aerococcaceae; g__Facklamia	Inf	11.01	Inf	2.69E-06	7.66E-06
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Leuconostocaceae; g__Weissella	Inf	2.12	Inf	5.22E-03	9.05E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Sarcina	Inf	3.82	Inf	4.58E-04	9.41E-04
	k__Bacteria; p__Fusobacteria; c__Fusobacteriia; o__Fusobacteriales; f__Leptotrichiaceae; g__Leptotrichia	Inf	3.82	Inf	4.58E-04	9.41E-04
	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Alcaligenaceae; g__Oligella	Inf	2.12	Inf	5.22E-03	9.05E-03
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__	Inf	3.82	Inf	4.58E-04	9.41E-04
	k__Archaea; p__Euryarchaeota; c__Methanobacteria; o__Methanobacteriales; f__Methanobacteriaceae; g__Methanobrevibacter	0	0	0.47	5.22E-03	9.05E-03
	k__Bacteria; p__Actinobacteria; c__Coriobacteriia; o__Coriobacteriales; f__Coriobacteriaceae; g__	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Actinobacteria; c__Coriobacteriia; o__Coriobacteriales; f__Coriobacteriaceae; g__Collinsella	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__; g__	0	0	0.19	1.08E-04	2.44E-04
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__[Odoribacteraceae]; g__Butyricimonas	0	0	0.47	5.22E-03	9.05E-03
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__[Paraprevotellaceae]; g__	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__[Paraprevotellaceae]; g__[Prevotella]	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__[Paraprevotellaceae]; g__CF231	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__[Paraprevotellaceae]; g__YRC22	0	0	0.26	4.58E-04	9.41E-04
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__p-2534-18B5; g__	0	0	0.35	1.65E-03	3.15E-03
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Porphyromonadaceae; g__Parabacteroides	0	0	0.09	2.69E-06	7.66E-06

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Chlamydiae; c__Chlamydiia; o__Chlamydiales; f__Chlamydiaceae; g__Chlamydia	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Cyanobacteria; c__4C0d-2; o__YS2; f__; g__	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Deferribacteres; c__Deferribacteres; o__Deferribacterales; f__Deferribacteraceae; g__Mucispirillum	0	0	0.26	4.58E-04	9.41E-04
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Mogibacteriaceae]; g__	0	0	0.19	1.08E-04	2.44E-04
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__	0	0	0.19	1.08E-04	2.44E-04
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__[Ruminococcus]	0	0	0.09	2.69E-06	7.66E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Anaerostipes	0	0	0.14	2.02E-05	4.98E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Blautia	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Butyrivibrio	0	0	0.14	2.02E-05	4.98E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Coprococcus	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Dorea	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Lachnobacterium	0	0	0.14	2.02E-05	4.98E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Lachnospira	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Roseburia	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Shuttleworthia	0	0	0.09	2.69E-06	7.66E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptococcaceae; g__Peptococcus	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Faecalibacterium	0	0	0.47	5.22E-03	9.05E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Oscillospira	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus	0	0	0.05	1.92E-07	7.11E-07

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Acidaminococcus	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Mitsuokella	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Selenomonas	0	0	0.14	2.02E-05	4.98E-05
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__	0	0	0.09	2.69E-06	7.66E-06
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__[Eubacterium]	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Bulleidia	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__Catenibacterium	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__p-75-a5	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__RFN20	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Planctomycetes; c__Planctomycetia; o__Pirellulales; f__Pirellulaceae; g__	0	0	0.09	2.69E-06	7.66E-06
	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__RF32; f__; g__	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Alcaligenaceae; g__Sutterella	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Tremblayales; f__; g__	0.041	0	0.37	1.20E-03	2.43E-03
	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Desulfovibrionales; f__Desulfovibrionaceae; g__	0	0	0.14	2.02E-05	4.98E-05
	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__Desulfovibrionales; f__Desulfovibrionaceae; g__Desulfovibrio	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__GMD14H09; f__; g__	0	0	0.09	2.69E-06	7.66E-06
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Aeromonadales; f__Succinivibrionaceae; g__	0	0	0.47	5.22E-03	9.05E-03
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Aeromonadales; f__Succinivibrionaceae; g__Anaerobiospirillum	0	0	0.14	2.02E-05	4.98E-05

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Aeromonadales; f__Succinivibrionaceae; g__Succinivibrio	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Spirochaetes; c__Spirochaetes; o__Sphaerochaetales; f__Sphaerochaetaceae; g__Sphaerochaeta	0	0	0.35	1.65E-03	3.15E-03
	k__Bacteria; p__Spirochaetes; c__Spirochaetes; o__Spirochaetales; f__Spirochaetaceae; g__Treponema	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__Tenericutes; c__Mollicutes; o__Anaeroplasmatales; f__Anaeroplasmataceae; g__	0	0	0.35	1.65E-03	3.15E-03
	k__Bacteria; p__Tenericutes; c__Mollicutes; o__RF39; f__; g__	0	0	0.05	1.92E-07	7.11E-07
	k__Bacteria; p__TM7; c__TM7-3; o__CW040; f__F16; g__	0	0	0.19	1.08E-04	2.44E-04

Proximal colon vs distal colon

	Genus	Odds Ratio	lower	upper	*p*-values	adjPvalues
	k__Archaea; p__Euryarchaeota; c__Methanobacteria; o__Methanobacteriales; f__Methanobacteriaceae; g__Methanosphaera	0	0	0.19	1.08E-04	1.36E-03
	k__Archaea; p__Euryarchaeota; c__Thermoplasmata; o__E2; f__[Methanomassiliicoccaceae]; g__vadinCA11	0	0	0.09	2.69E-06	6.80E-05
	k__Bacteria; p__Elusimicrobia; c__Elusimicrobia; o__Elusimicrobiales; f__Elusimicrobiaceae; g__	0	0	0.26	4.58E-04	5.14E-03
	k__Bacteria; p__Fibrobacteres; c__Fibrobacteria; o__Fibrobacterales; f__Fibrobacteraceae; g__Fibrobacter	0	0	0.09	2.69E-06	6.80E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Mogibacteriaceae]; g__Anaerovorax	0	0	0.14	2.02E-05	4.08E-04
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__L7A_E11	0	0	0.19	1.08E-04	1.36E-03
	k__Bacteria; p__Synergistetes; c__Synergistia; o__Synergistales; f__Dethiosulfovibrionaceae; g__	0	0	0.09	2.69E-06	6.80E-05
	k__Bacteria; p__Tenericutes; c__RF3; o__ML615J-28; f__; g__	0	0	0.19	1.08E-04	1.36E-03
	k__Bacteria; p__Verrucomicrobia; c__Verruco-5; o__WCHB1-41; f__RFP12; g__	0	0	0.09	2.69E-06	6.80E-05

Supplementary Table S2. This table contains the results of the differential abundance analysis at the genus level performed with metagenomeSeq[24]. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which genus is more abundant in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and, purple, distal colon.

Duodenum vs jenunum

	Genus	*p*-values	adjPvalues
	k__Bacteria; p__Cyanobacteria; c__Chloroplast; o__Streptophyta; f__; g__	4.45E-07	6.45E-06
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Aerococcaceae; g__Facklamia	1.69E-11	1.47E-09
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Leuconostocaceae; g__Weissella	1.13E-04	9.85E-04
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Moraxella	1.28E-06	1.59E-05
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Psychrobacter	3.87E-05	3.74E-04
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Turicibacterales; f__Turicibacteraceae; g__Turicibacter	1.56E-09	4.53E-08
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__	8.48E-04	6.15E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__SMB53	2.42E-09	5.27E-08
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptostreptococcaceae; g__	1.36E-09	4.53E-08
	k__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Helicobacter	2.39E-08	4.16E-07
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae; g__	1.85E-05	2.01E-04
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__Actinobacillus	3.91E-04	3.09E-03

Jejunum vs ileum

	Genus	*p*-values	adjPvalues
	k__Bacteria; p__Cyanobacteria; c__Chloroplast; o__Streptophyta; f__; g__	4.87E-06	7.06E-05
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus	2.20E-05	2.13E-04
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Turicibacterales; f__Turicibacteraceae; g__Turicibacter	8.46E-04	6.13E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__	2.85E-07	8.27E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__	9.74E-07	1.88E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__SMB53	1.39E-07	8.03E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptostreptococcaceae; g__	8.43E-04	6.13E-03
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae; g__	7.42E-06	8.61E-05
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__Actinobacillus	1.09E-03	7.02E-03

Ileum vs proximal colon

	Genus	*p*-values	adjPvalues
	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Corynebacteriaceae; g__Corynebacterium	1.33E-06	3.13E-06
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus	1.23E-05	2.76E-05
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus	4.56E-11	1.53E-10
	k__Bacteria; p__Firmicutes; c__Bacilli; o__Turicibacterales; f__Turicibacteraceae; g__Turicibacter	9.68E-16	6.50E-15
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__	2.31E-11	8.36E-11
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__Clostridium	3.82E-13	1.79E-12
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__SMB53	8.36E-12	3.27E-11
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Peptostreptococcaceae; g__	2.86E-08	7.90E-08
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Veillonella	1.95E-08	5.74E-08
	k__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Flexispira	4.57E-08	1.19E-07
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae; g__	8.85E-13	3.78E-12
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__Actinobacillus	8.93E-15	5.24E-14
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__Aggregatibacter	1.90E-08	5.74E-08
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Psychrobacter	8.85E-07	2.19E-06
	k__Bacteria; p__Tenericutes; c__Mollicutes; o__Mycoplasmatales; f__Mycoplasmataceae; g__Mycoplasma	1.73E-03	3.70E-03
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae; g__Prevotella	1.85E-36	8.71E-35
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__; g__	7.66E-14	4.00E-13
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__	1.53E-17	1.20E-16
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Anaerovibrio	5.04E-34	1.19E-32
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Dialister	2.08E-21	1.95E-20
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Megasphaera	1.11E-30	1.74E-29
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Phascolarctobacterium	4.34E-25	5.10E-24

Proximal colon vs distal colon

	Genus	*p*-values	adjPvalues
	k__Bacteria; p__Cyanobacteria; c__4C0d-2; o__YS2; f__; g__	2.41E-03	7.65E-03
	k__Bacteria; p__Deferribacteres; c__Deferribacteres; o__Deferribacterales; f__Deferribacteraceae; g__Mucispirillum	2.50E-07	1.44E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__	2.73E-03	8.36E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Clostridiaceae; g__SMB53	8.83E-07	4.78E-06
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Acidaminococcus	7.55E-05	2.89E-04
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Dialister	1.58E-04	5.61E-04
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Megasphaera	9.96E-06	4.36E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Mitsuokella	2.14E-05	8.57E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Veillonellaceae; g__Veillonella	5.31E-10	4.44E-09
	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Tremblayales; f__; g__	8.37E-11	9.63E-10
	k__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Campylobacteraceae; g__Campylobacter	1.33E-05	5.55E-05
	k__Bacteria; p__Proteobacteria; c__Epsilonproteobacteria; o__Campylobacterales; f__Helicobacteraceae; g__Helicobacter	1.34E-10	1.37E-09
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae; g__	1.97E-10	1.81E-09
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__Actinobacillus	1.17E-20	3.15E-19
	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pasteurellales; f__Pasteurellaceae; g__Aggregatibacter	1.50E-07	9.87E-07

	Genus	*p*-values	adjPvalues
	k__Archaea; p__Euryarchaeota; c__Methanobacteria; o__Methanobacteriales; f__Methanobacteriaceae; g__Methanobrevibacter	6.29E-10	4.82E-09
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__; g__	7.45E-06	3.43E-05
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__p-2534-18B5; g__	1.89E-07	1.16E-06
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Porphyromonadaceae; g__Parabacteroides	4.28E-04	1.46E-03
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__RF16; g__	3.03E-06	1.47E-05
	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__S24-7; g__	3.08E-03	9.15E-03
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Mogibacteriaceae]; g__	2.89E-06	1.47E-05
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Christensenellaceae; g__	1.83E-22	1.68E-20
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Lachnospiraceae; g__Shuttleworthia	1.08E-04	3.97E-04
	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Ruminococcaceae; g__Ruminococcus	1.60E-03	5.25E-03
	k__Bacteria; p__Firmicutes; c__Erysipelotrichi; o__Erysipelotrichales; f__Erysipelotrichaceae; g__RFN20	2.31E-16	4.26E-15
	k__Bacteria; p__Planctomycetes; c__Planctomycetia; o__Pirellulales; f__Pirellulaceae; g__	1.37E-20	3.15E-19
	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__GMD14H09; f__; g__	6.57E-16	1.01E-14
	k__Bacteria; p__Spirochaetes; c__Spirochaetes; o__Sphaerochaetales; f__Sphaerochaetaceae; g__Sphaerochaeta	3.66E-21	1.68E-19
	k__Bacteria; p__Spirochaetes; c__Spirochaetes; o__Spirochaetales; f__Spirochaetaceae; g__Treponema	2.80E-12	3.68E-11
	k__Bacteria; p__WPS-2; c__; o__; f__; g__	5.92E-09	4.19E-08

Supplementary Table S3. This table contains the results of the differential abundance analysis performed with DESeq2[27] for the KEGG[26] orthologies (KOs) predicted with PICRUSt[25] at the pathway level. Each sheet represents one of the four consecutive correlations between the five sections for the 13 pigs. The colour shows which pathway is more abundant in that section when comparing the two sections: red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon.

Duodenum vs jenunum

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	N-Glycan biosynthesis	2934.32	-1.34	0.18	-7.31	2.74E-13	7.29E-11
	Photosynthesis - antenna proteins	9285.34	-1.33	0.21	-6.36	1.96E-10	2.60E-08
	Carotenoid biosynthesis	4769.17	-1.27	0.21	-6.08	1.18E-09	7.82E-08
	Flavonoid biosynthesis	1714.87	-1.23	0.21	-5.91	3.51E-09	1.87E-07
	Steroid biosynthesis	1265.53	-1.22	0.21	-5.84	5.32E-09	2.36E-07
	Meiosis - yeast	2441.89	-1.21	0.21	-5.78	7.30E-09	2.77E-07
	Calcium signaling pathway	546.28	-1.09	0.21	-5.31	1.09E-07	2.91E-06
	Fluorobenzoate degradation	1536.4	-1.1	0.21	-5.28	1.27E-07	3.06E-06
	Chlorocyclohexane and chlorobenzene degradation	3647.5	-0.99	0.2	-4.85	1.26E-06	2.40E-05
	Parkinsons disease	3821.63	-0.97	0.21	-4.72	2.39E-06	4.24E-05
	p53 signaling pathway	464.32	-0.95	0.21	-4.63	3.62E-06	6.02E-05
	Colorectal cancer	444.02	-0.94	0.21	-4.56	5.14E-06	6.50E-05
	Influenza A	444.02	-0.94	0.21	-4.56	5.14E-06	6.50E-05
	Small cell lung cancer	444.02	-0.94	0.21	-4.56	5.14E-06	6.50E-05
	Toxoplasmosis	444.02	-0.94	0.21	-4.56	5.14E-06	6.50E-05
	Viral myocarditis	444.02	-0.94	0.21	-4.56	5.14E-06	6.50E-05
	Protein processing in endoplasmic reticulum	12069.88	-0.46	0.1	-4.55	5.43E-06	6.57E-05
	Protein digestion and absorption	265.24	-0.91	0.21	-4.42	9.68E-06	1.07E-04
	Porphyrin and chlorophyll metabolism	230785.57	-0.37	0.09	-4.13	3.56E-05	2.96E-04
	Cardiac muscle contraction	3377.61	-0.85	0.21	-4.11	3.99E-05	3.22E-04
	Metabolism of xenobiotics by cytochrome P450	20811.98	-0.62	0.16	-3.97	7.06E-05	5.37E-04
	Photosynthesis proteins	144911.7	-0.44	0.12	-3.76	1.68E-04	1.15E-03
	beta-Alanine metabolism	38310.16	-0.27	0.07	-3.72	1.99E-04	1.32E-03
	Drug metabolism - cytochrome P450	23390.77	-0.53	0.15	-3.52	4.24E-04	2.68E-03
	Photosynthesis	134119.19	-0.37	0.11	-3.46	5.45E-04	3.32E-03
	Lipoic acid metabolism	9288.07	-0.6	0.18	-3.37	7.61E-04	4.22E-03
	Phosphatidylinositol signaling system	26856.31	-0.26	0.08	-3.25	1.15E-03	6.12E-03
	Sulfur metabolism	57421.49	-0.15	0.05	-3.11	1.88E-03	9.64E-03

Saa

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Transporters	1794864.39	0.29	0.05	6.15	7.90E-10	7.01E-08
	ABC transporters	845546.79	0.26	0.05	5.61	2.07E-08	6.89E-07
	Fructose and mannose metabolism	226727.63	0.33	0.06	5.43	5.77E-08	1.71E-06
	Amino sugar and nucleotide sugar metabolism	405302.7	0.26	0.05	5.04	4.56E-07	1.01E-05
	G protein-coupled receptors	136.33	0.9	0.18	5	5.84E-07	1.20E-05
	Translation proteins	227053.85	0.14	0.03	4.53	5.82E-06	6.74E-05
	Drug metabolism - other enzymes	85188.88	0.35	0.08	4.37	1.25E-05	1.33E-04
	Transcription machinery	205525.95	0.22	0.05	4.35	1.33E-05	1.36E-04
	Biosynthesis of ansamycins	19808.89	0.53	0.12	4.3	1.72E-05	1.70E-04
	Energy metabolism	159544.66	0.4	0.09	4.28	1.85E-05	1.76E-04
	Methane metabolism	268988.84	0.12	0.03	4.17	3.10E-05	2.84E-04
	DNA replication proteins	329218.49	0.16	0.04	4.16	3.22E-05	2.85E-04
	Phosphotransferase system (PTS)	296296.2	0.64	0.16	4.14	3.42E-05	2.93E-04
	Transcription factors	477950.16	0.31	0.08	4.05	5.10E-05	3.99E-04
	Chromosome	417173.84	0.12	0.03	3.89	1.00E-04	7.42E-04
	General function prediction only	905010.41	0.07	0.02	3.85	1.20E-04	8.65E-04
	Prion diseases	885.21	0.78	0.21	3.79	1.52E-04	1.06E-03
	Ribosome Biogenesis	377704.46	0.16	0.04	3.63	2.87E-04	1.86E-03
	Bacterial toxins	43190.11	0.31	0.09	3.45	5.62E-04	3.32E-03
	Nitrotoluene degradation	7338.68	0.64	0.18	3.45	5.54E-04	3.32E-03
	Pentose and glucuronate interconversions	107251.35	0.3	0.09	3.4	6.68E-04	3.86E-03
	Function unknown	406540.31	0.18	0.05	3.39	7.02E-04	3.97E-03
	Germination	4472.68	0.68	0.21	3.29	9.86E-04	5.35E-03
	Galactose metabolism	216679.76	0.35	0.11	3.16	1.57E-03	8.18E-03

Jejunum vs ileum

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Basal transcription factors	1931.69	-1.29	0.2	-6.54	6.03E-11	7.87E-09
	Flavonoid biosynthesis	276.03	-1.14	0.2	-5.72	1.06E-08	5.54E-07
	Drug metabolism - cytochrome P450	15439.71	-0.86	0.15	-5.69	1.28E-08	5.55E-07
	Retinol metabolism	10417.12	-0.93	0.18	-5.25	1.51E-07	3.95E-06
	Prenyltransferases	102942.58	-0.29	0.06	-5.23	1.69E-07	4.02E-06
	Stilbenoid, diarylheptanoid and gingerol biosynthesis	2215.25	-1.03	0.2	-5.2	2.01E-07	4.38E-06
	Mineral absorption	2278.05	-1.03	0.2	-5.13	2.86E-07	5.75E-06
	Glycosphingolipid biosynthesis - ganglio series	3868.98	-1.02	0.2	-5.11	3.26E-07	6.07E-06
	Steroid hormone biosynthesis	2029.52	-1.01	0.2	-5.07	3.99E-07	6.30E-06
	Carotenoid biosynthesis	918.61	-1.01	0.2	-5.06	4.10E-07	6.30E-06
	Glycolysis / Gluconeogenesis	377156.88	-0.23	0.04	-5.06	4.11E-07	6.30E-06
	Amoebiasis	2042.74	-1	0.2	-5.01	5.46E-07	7.92E-06
	Primary bile acid biosynthesis	6086.65	-0.9	0.18	-4.89	1.02E-06	1.40E-05
	Calcium signaling pathway	83.66	-0.87	0.18	-4.8	1.55E-06	1.92E-05
	Photosynthesis - antenna proteins	1421.08	-0.86	0.18	-4.77	1.82E-06	2.07E-05
	Meiosis - yeast	471.01	-0.93	0.2	-4.74	2.18E-06	2.28E-05
	Metabolism of xenobiotics by cytochrome P450	13218.83	-0.75	0.16	-4.74	2.13E-06	2.28E-05
	Biosynthesis of siderophore group nonribosomal peptides	6773.48	-0.56	0.12	-4.68	2.81E-06	2.82E-05
	Lysosome	16366.19	-0.79	0.17	-4.58	4.70E-06	4.47E-05
	Glycerolipid metabolism	137035.29	-0.2	0.04	-4.57	4.79E-06	4.47E-05
	Restriction enzyme	67126.65	-0.48	0.11	-4.52	6.12E-06	5.51E-05
	Secondary bile acid biosynthesis	6002.75	-0.85	0.19	-4.49	7.15E-06	6.22E-05
	Glycosaminoglycan degradation	11666.87	-0.76	0.17	-4.41	1.05E-05	8.59E-05
	Primary immunodeficiency	17334.41	-0.5	0.11	-4.36	1.30E-05	1.03E-04
	beta-Lactam resistance	7758.88	-0.57	0.14	-4.1	4.07E-05	2.87E-04
	Steroid biosynthesis	262.36	-0.77	0.19	-4.05	5.18E-05	3.56E-04
	Type II diabetes mellitus	18827.47	-0.31	0.08	-3.9	9.65E-05	6.30E-04
	Nucleotide excision repair	129349.42	-0.22	0.06	-3.84	1.21E-04	7.72E-04
	Taurine and hypotaurine metabolism	36274.09	-0.18	0.05	-3.66	2.49E-04	1.51E-03
	Glutathione metabolism	84601.1	-0.14	0.04	-3.64	2.76E-04	1.64E-03
	Function unknown	469302.42	-0.14	0.04	-3.6	3.14E-04	1.71E-03
	Others	322021.24	-0.22	0.06	-3.58	3.38E-04	1.80E-03
	Staphylococcus aureus infection	29004.77	-0.6	0.17	-3.52	4.26E-04	2.10E-03
	Bacterial toxins	48807.58	-0.27	0.08	-3.41	6.40E-04	2.97E-03
	Xylene degradation	23672.56	-0.32	0.09	-3.4	6.69E-04	2.98E-03

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Chloroalkane and chloroalkene degradation	56947.83	-0.35	0.1	-3.37	7.45E-04	3.15E-03
	Purine metabolism	745806.31	-0.1	0.03	-3.38	7.31E-04	3.15E-03
	Pentose phosphate pathway	259171.05	-0.14	0.04	-3.33	8.66E-04	3.50E-03
	Other ion-coupled transporters	415651.25	-0.17	0.05	-3.3	9.56E-04	3.78E-03
	Penicillin and cephalosporin biosynthesis	12912.65	-0.52	0.16	-3.23	1.26E-03	4.76E-03
	Colorectal cancer	102.1	-0.59	0.19	-3.19	1.40E-03	4.93E-03
	Influenza A	102.1	-0.59	0.19	-3.19	1.40E-03	4.93E-03
	Small cell lung cancer	102.1	-0.59	0.19	-3.19	1.40E-03	4.93E-03
	Toxoplasmosis	102.1	-0.59	0.19	-3.19	1.40E-03	4.93E-03
	Viral myocarditis	102.1	-0.59	0.19	-3.19	1.40E-03	4.93E-03
	Fluorobenzoate degradation	447.53	-0.63	0.2	-3.17	1.51E-03	5.17E-03
	Dioxin degradation	28328.92	-0.28	0.09	-3.12	1.80E-03	6.11E-03
	Bile secretion	30.66	-0.55	0.18	-3.02	2.56E-03	8.35E-03
	Systemic lupus erythematosus	30.66	-0.55	0.18	-3.02	2.56E-03	8.35E-03
	Transcription related proteins	5689.83	-0.6	0.2	-3	2.67E-03	8.60E-03
	Tetracycline biosynthesis	54319.9	0.33	0.04	7.62	2.58E-14	6.72E-12
	Polyketide sugar unit biosynthesis	39058.59	0.35	0.06	6.22	5.04E-10	4.38E-08
	Histidine metabolism	130434.25	0.47	0.08	5.76	8.22E-09	5.36E-07
	Fatty acid biosynthesis	152502.97	0.21	0.04	5.59	2.27E-08	8.47E-07
	Lipid biosynthesis proteins	160586.14	0.14	0.03	5.41	6.20E-08	1.96E-06
	Ethylbenzene degradation	16680.52	0.47	0.09	5.4	6.75E-08	1.96E-06
	Glycine, serine and threonine metabolism	208892.63	0.33	0.07	4.81	1.48E-06	1.92E-05
	Cysteine and methionine metabolism	282941.51	0.19	0.04	4.79	1.66E-06	1.97E-05
	Sulfur metabolism	65921.23	0.19	0.04	4.45	8.60E-06	7.24E-05
	Translation proteins	284354.11	0.13	0.03	4.29	1.75E-05	1.35E-04
	Lysine biosynthesis	215479.33	0.14	0.03	4.22	2.44E-05	1.82E-04
	Pyruvate metabolism	327232.93	0.05	0.01	4.17	3.11E-05	2.25E-04

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Phosphonate and phosphinate metabolism	20497.97	0.35	0.09	3.92	8.95E-05	5.99E-04
	Biosynthesis of unsaturated fatty acids	37202.27	0.27	0.07	3.7	2.13E-04	1.32E-03
	Vibrio cholerae pathogenic cycle	28979.07	0.34	0.09	3.62	2.98E-04	1.69E-03
	Vitamin B6 metabolism	40986.86	0.47	0.13	3.62	2.96E-04	1.69E-03
	Novobiocin biosynthesis	33516.46	0.54	0.15	3.61	3.05E-04	1.69E-03
	Bacterial secretion system	172773.08	0.16	0.05	3.55	3.79E-04	1.98E-03
	Biosynthesis of vancomycin group antibiotics	12859.86	0.18	0.05	3.54	3.95E-04	1.98E-03
	Styrene degradation	4552.96	0.71	0.2	3.55	3.88E-04	1.98E-03
	RIG-I-like receptor signaling pathway	1861.15	0.69	0.2	3.47	5.18E-04	2.50E-03
	Ether lipid metabolism	1748.8	0.67	0.2	3.43	6.08E-04	2.89E-03
	Valine, leucine and isoleucine biosynthesis	192109.35	0.34	0.1	3.41	6.49E-04	2.97E-03
	Pantothenate and CoA biosynthesis	167234.68	0.16	0.05	3.4	6.73E-04	2.98E-03
	General function prediction only	1075183.11	0.05	0.02	3.37	7.49E-04	3.15E-03
	Tropane, piperidine and pyridine alkaloid biosynthesis	29061.57	0.48	0.14	3.35	8.11E-04	3.36E-03
	Peroxisome	45595.61	0.33	0.1	3.33	8.73E-04	3.50E-03
	Lysine degradation	42914.49	0.29	0.09	3.28	1.05E-03	4.11E-03
	Butanoate metabolism	217341.04	0.11	0.03	3.26	1.13E-03	4.33E-03
	C5-Branched dibasic acid metabolism	71162.7	0.43	0.14	3.19	1.42E-03	4.93E-03
	Phenylalanine, tyrosine and tryptophan biosynthesis	165948.77	0.43	0.14	3.07	2.15E-03	7.19E-03
	One carbon pool by folate	163257.45	0.1	0.03	2.98	2.84E-03	9.03E-03

Ileum vs proximal colon

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Transcription factors	577051.78	-0.72	0.03	-24	5.92E-127	1.59E-124
	Transporters	2142011.7	-0.63	0.03	-23.3	2.57E-120	3.45E-118
	Pyruvate metabolism	343513.73	-0.44	0.02	-22.4	2.47E-111	2.21E-109
	Tetracycline biosynthesis	50195.23	-1.33	0.07	-19.9	7.99E-88	5.36E-86
	ABC transporters	1097788.77	-0.56	0.03	-19.8	6.46E-87	3.46E-85
	Signal transduction mechanisms	162689.36	-0.52	0.03	-17.4	7.93E-68	2.66E-66
	Dioxin degradation	19570.75	-1.71	0.1	-16.5	5.92E-61	1.59E-59
	Synthesis and degradation of ketone bodies	17624.35	-1.44	0.09	-16.4	1.00E-60	2.44E-59
	Benzoate degradation	81189.03	-0.88	0.05	-16.3	2.19E-59	4.89E-58
	Xylene degradation	15967.47	-1.74	0.11	-16	2.78E-57	5.73E-56
	Propanoate metabolism	183130.67	-0.41	0.03	-15.8	6.15E-56	1.18E-54
	Others	302154.06	-0.46	0.03	-15.5	2.13E-54	3.81E-53
	Phosphotransferase system (PTS)	228498.57	-1.84	0.12	-15	4.81E-51	8.05E-50
	Atrazine degradation	6426.64	-1.7	0.12	-14.7	4.22E-49	6.65E-48
	Phosphonate and phosphinate metabolism	20156.36	-1.05	0.07	-14.2	5.44E-46	7.29E-45
	Sulfur relay system	96351.89	-0.76	0.06	-13.4	1.08E-40	1.21E-39
	Chagas disease (American trypanosomiasis)	442.64	-3.8	0.29	-13	1.44E-38	1.43E-37
	Glycerolipid metabolism	128338.38	-0.51	0.04	-12.2	3.44E-34	3.18E-33
	Fatty acid biosynthesis	170960.61	-0.39	0.03	-11.9	8.23E-33	7.36E-32
	Staphylococcus aureus infection	14788.44	-2.69	0.24	-11.2	5.31E-29	4.45E-28
	Fatty acid metabolism	78978.94	-0.54	0.05	-10.9	1.75E-27	1.38E-26
	Apoptosis	2373.76	-2.73	0.25	-10.8	5.27E-27	4.04E-26
	Naphthalene degradation	44269.61	-0.52	0.05	-10.7	1.69E-26	1.26E-25
	Butanoate metabolism	229879.26	-0.47	0.05	-10.4	3.62E-25	2.62E-24
	Secretion system	459469.28	-0.36	0.04	-9.65	4.80E-22	2.99E-21
	Tyrosine metabolism	117847.99	-0.35	0.04	-9.58	9.78E-22	5.82E-21
	Proximal tubule bicarbonate reclamation	5963.19	-1.72	0.18	-9.39	5.98E-21	3.48E-20
	Replication, recombination and repair proteins	306824.72	-0.4	0.04	-9.35	8.48E-21	4.84E-20
	Bacterial toxins	40897.12	-0.73	0.08	-8.99	2.53E-19	1.38E-18
	D-Alanine metabolism	46313.68	-0.37	0.04	-8.83	1.06E-18	5.56E-18
	Metabolism of cofactors and vitamins	51206.83	-0.55	0.06	-8.73	2.46E-18	1.25E-17
	Transcription related proteins	2495.27	-2.2	0.26	-8.57	1.00E-17	4.89E-17
	Nucleotide metabolism	17612.72	-0.67	0.08	-8.52	1.56E-17	7.44E-17
	Tryptophan metabolism	50652.26	-0.58	0.07	-8.51	1.67E-17	7.87E-17
	Huntingtons disease	15197.08	-1.16	0.14	-8.5	1.85E-17	8.54E-17
	Ion channels	9913.93	-1.54	0.18	-8.48	2.26E-17	1.03E-16

Dssd

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Chloroalkane and chloroalkene degradation	46998.77	-0.69	0.08	-8.23	1.80E-16	7.91E-16
	Function unknown	460708.33	-0.43	0.05	-8.21	2.12E-16	9.19E-16
	G protein-coupled receptors	223.64	-2.98	0.37	-7.96	1.68E-15	7.11E-15
	Protein kinases	85915.73	-0.4	0.05	-7.67	1.76E-14	7.03E-14
	RIG-I-like receptor signaling pathway	1876.81	-2.43	0.33	-7.48	7.47E-14	2.86E-13
	African trypanosomiasis	492.15	-2.45	0.33	-7.38	1.56E-13	5.88E-13
	Aminobenzoate degradation	45809.55	-0.46	0.06	-7.28	3.27E-13	1.20E-12
	Glycerophospholipid metabolism	184164.39	-0.21	0.03	-7.1	1.21E-12	4.31E-12
	Pentose phosphate pathway	273495.89	-0.19	0.03	-7	2.52E-12	8.90E-12
	Two-component system	494263.47	-0.49	0.07	-6.94	3.96E-12	1.38E-11
	Ether lipid metabolism	1830.24	-2.22	0.33	-6.71	1.93E-11	6.54E-11
	Ethylbenzene degradation	19002.33	-0.59	0.09	-6.46	1.02E-10	3.36E-10
	Ubiquitin system	1913.57	-1.48	0.23	-6.38	1.77E-10	5.71E-10
	RNA transport	45338.9	-0.42	0.07	-6.13	8.84E-10	2.69E-09
	Glycolysis / Gluconeogenesis	378927.26	-0.24	0.04	-6.1	1.05E-09	3.13E-09
	Limonene and pinene degradation	32051.14	-0.52	0.09	-5.87	4.26E-09	1.23E-08
	Caprolactam degradation	10037.42	-1.36	0.23	-5.82	5.97E-09	1.70E-08
	Base excision repair	159747.94	-0.13	0.02	-5.73	1.03E-08	2.91E-08
	Lysine degradation	47179.29	-0.51	0.09	-5.65	1.63E-08	4.45E-08
	Pentose and glucuronate interconversions	149213.76	-0.29	0.05	-5.39	7.13E-08	1.84E-07
	RNA polymerase	67180.51	-0.22	0.04	-5.36	8.14E-08	2.08E-07
	Bacterial invasion of epithelial cells	538.02	-1.84	0.34	-5.35	9.03E-08	2.28E-07
	Prion diseases	1937.33	-1.42	0.27	-5.17	2.32E-07	5.82E-07
	Amino acid metabolism	70899.32	-0.39	0.08	-5.11	3.22E-07	8.00E-07
	Flavone and flavonol biosynthesis	3714.2	-1.16	0.24	-4.93	8.16E-07	2.01E-06
	Inorganic ion transport and metabolism	61759.77	-0.45	0.09	-4.92	8.86E-07	2.15E-06
	General function prediction only	1267585.65	-0.07	0.01	-4.9	9.65E-07	2.29E-06
	Drug metabolism - cytochrome P450	9007.17	-0.97	0.2	-4.82	1.44E-06	3.35E-06
	Penicillin and cephalosporin biosynthesis	8588.57	-1.03	0.21	-4.79	1.68E-06	3.88E-06
	Glycosaminoglycan biosynthesis - chondroitin sulfate	15.95	-1.93	0.4	-4.77	1.89E-06	4.32E-06
	Biosynthesis and biodegradation of secondary metabolites	14328.91	-0.52	0.11	-4.72	2.42E-06	5.44E-06
	Amyotrophic lateral sclerosis (ALS)	4236.84	-1.12	0.25	-4.55	5.48E-06	1.19E-05
	Galactose metabolism	239176.25	-0.29	0.06	-4.49	7.21E-06	1.55E-05
	Valine, leucine and isoleucine biosynthesis	235987.82	-0.23	0.05	-4.41	1.02E-05	2.17E-05
	Metabolism of xenobiotics by cytochrome P450	8430.76	-0.79	0.2	-3.92	9.00E-05	1.80E-04

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Biosynthesis of unsaturated fatty acids	43838.04	-0.28	0.08	-3.7	2.18E-04	4.26E-04
	Glutathione metabolism	88813.08	-0.22	0.06	-3.66	2.50E-04	4.86E-04
	Phosphatidylinositol signaling system	29128.39	-0.34	0.09	-3.59	3.30E-04	6.32E-04
	Photosynthesis - antenna proteins	96.17	-1.36	0.38	-3.58	3.39E-04	6.45E-04
	Ribosome Biogenesis	535919.81	-0.11	0.03	-3.46	5.42E-04	1.00E-03
	Fructose and mannose metabolism	305013.75	-0.11	0.03	-3.34	8.29E-04	1.51E-03
	Vibrio cholerae pathogenic cycle	35383.16	-0.23	0.07	-3.19	1.43E-03	2.57E-03
	MAPK signaling pathway - yeast	12652.65	-0.26	0.08	-3.17	1.54E-03	2.72E-03
	Valine, leucine and isoleucine degradation	85211.9	-0.18	0.06	-3.13	1.75E-03	3.07E-03
	Inositol phosphate metabolism	29623.66	-0.39	0.13	-3.12	1.79E-03	3.11E-03
	Lipoic acid metabolism	8749.45	-0.65	0.22	-3.02	2.53E-03	4.32E-03
	Endocytosis	24.18	-1.19	0.4	-2.96	3.08E-03	5.12E-03
	Fc gamma R-mediated phagocytosis	24.18	-1.19	0.4	-2.96	3.08E-03	5.12E-03
	GnRH signaling pathway	24.18	-1.19	0.4	-2.96	3.08E-03	5.12E-03
	Pertussis	7739.37	-0.95	0.33	-2.85	4.33E-03	7.07E-03
	Bacterial chemotaxis	193481.63	-0.39	0.14	-2.85	4.30E-03	7.07E-03
	alpha-Linolenic acid metabolism	2798.73	-0.82	0.29	-2.84	4.45E-03	7.22E-03
	Basal transcription factors	549.42	-1.07	0.38	-2.81	4.91E-03	7.93E-03
	Carbon fixation in photosynthetic organisms	216092.45	0.26	0.01	19.5	6.06E-85	2.71E-83
	Biosynthesis of vancomycin group antibiotics	21955.64	0.78	0.04	18.9	2.10E-79	8.06E-78
	Protein digestion and absorption	7996.96	5.62	0.34	16.7	2.45E-62	7.29E-61
	Other glycan degradation	56226.57	1.51	0.1	14.6	2.90E-48	4.31E-47
	One carbon pool by folate	230634.11	0.38	0.03	14.5	1.46E-47	2.06E-46
	Zeatin biosynthesis	21529.39	0.62	0.04	14	1.34E-44	1.66E-43
	Cellular antigens	15178.28	2.55	0.18	14	1.36E-44	1.66E-43
	Protein processing in endoplasmic reticulum	22088.4	1.34	0.1	13.5	1.15E-41	1.34E-40
	Streptomycin biosynthesis	100939.74	0.46	0.03	13.3	3.32E-40	3.56E-39
	N-Glycan biosynthesis	6865.26	3.43	0.26	13	1.34E-38	1.39E-37
	Glycosphingolipid biosynthesis - globo series	28438.84	1.27	0.1	12.2	2.53E-34	2.42E-33
	Cyanoamino acid metabolism	83977.72	0.48	0.04	11.6	4.94E-31	4.27E-30
	Polyketide sugar unit biosynthesis	65224.39	0.6	0.05	11.2	5.73E-29	4.66E-28
	Adipocytokine signaling pathway	18079.09	1.24	0.12	9.95	2.57E-23	1.82E-22
	Meiosis - yeast	1223.87	3.22	0.32	9.92	3.49E-23	2.40E-22
	Sphingolipid metabolism	47345.78	0.78	0.08	9.81	9.90E-23	6.64E-22
	Phenylpropanoid biosynthesis	33004.3	1.12	0.11	9.8	1.16E-22	7.61E-22
	Oxidative phosphorylation	384811.58	0.42	0.04	9.71	2.75E-22	1.76E-21
	Cell cycle - Caulobacter	188696.23	0.24	0.03	9.63	5.83E-22	3.55E-21

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Epithelial cell signaling in Helicobacter pylori infection	28694.2	0.59	0.06	9.29	1.50E-20	8.40E-20
	Nicotinate and nicotinamide metabolism	147738.27	0.3	0.03	8.87	7.60E-19	4.07E-18
	Glycosphingolipid biosynthesis - ganglio series	11696.31	2.89	0.33	8.78	1.66E-18	8.57E-18
	Cell division	21825.9	0.79	0.09	8.73	2.52E-18	1.25E-17
	Arachidonic acid metabolism	19680.7	0.92	0.11	8.44	3.07E-17	1.37E-16
	beta-Alanine metabolism	61558.04	0.64	0.08	7.96	1.70E-15	7.11E-15
	Prenyltransferases	124555.16	0.35	0.04	7.88	3.26E-15	1.34E-14
	Citrate cycle (TCA cycle)	209391.59	0.46	0.06	7.78	7.45E-15	3.03E-14
	Vitamin B6 metabolism	67287.39	0.46	0.06	7.6	3.05E-14	1.20E-13
	Flavonoid biosynthesis	275.04	2.15	0.28	7.54	4.53E-14	1.76E-13
	Chaperones and folding catalysts	374612.55	0.22	0.03	7.28	3.23E-13	1.20E-12
	Carbon fixation pathways in prokaryotes	355478.24	0.19	0.03	7.26	3.92E-13	1.42E-12
	Phenylalanine, tyrosine and tryptophan biosynthesis	258959.85	0.37	0.05	6.87	6.61E-12	2.27E-11
	Glycosaminoglycan degradation	16859.23	1.5	0.23	6.6	4.25E-11	1.42E-10
	Lysosome	22611.01	1.48	0.23	6.42	1.38E-10	4.52E-10
	Photosynthesis	148308.71	0.2	0.03	6.38	1.81E-10	5.77E-10
	PPAR signaling pathway	30456.12	0.56	0.09	6.22	4.87E-10	1.54E-09
	Protein folding and associated processing	215338.24	0.2	0.03	6.16	7.08E-10	2.21E-09
	Mineral absorption	3546.93	2.06	0.34	6.16	7.42E-10	2.29E-09
	NOD-like receptor signaling pathway	13866.99	0.79	0.13	6.11	1.00E-09	3.02E-09
	Biotin metabolism	48447.58	0.51	0.08	6.04	1.52E-09	4.48E-09
	Steroid hormone biosynthesis	3463.43	2.1	0.36	5.91	3.44E-09	1.00E-08
	Alanine, aspartate and glutamate metabolism	353825.03	0.13	0.02	5.67	1.40E-08	3.91E-08
	Proteasome	13447.46	0.73	0.13	5.66	1.53E-08	4.23E-08
	Lipopolysaccharide biosynthesis	117239.21	1.22	0.22	5.63	1.84E-08	4.99E-08
	Prostate cancer	13371.37	0.73	0.13	5.61	2.01E-08	5.38E-08
	Antigen processing and presentation	13368.85	0.73	0.13	5.61	2.08E-08	5.47E-08
	Progesterone-mediated oocyte maturation	13368.85	0.73	0.13	5.61	2.08E-08	5.47E-08
	Photosynthesis proteins	150520.73	0.18	0.03	5.4	6.75E-08	1.76E-07
	Membrane and intracellular structural molecules	205148.06	0.49	0.1	4.91	8.89E-07	2.15E-06
	Translation factors	206689.86	0.12	0.02	4.91	9.13E-07	2.19E-06
	Amino acid related enzymes	547831.61	0.09	0.02	4.84	1.28E-06	3.02E-06
	Pathways in cancer	19245.82	0.35	0.07	4.74	2.12E-06	4.81E-06
	Insulin signaling pathway	22448.3	0.35	0.08	4.61	4.03E-06	9.00E-06
	Lipopolysaccharide biosynthesis proteins	154484.52	0.8	0.18	4.59	4.44E-06	9.82E-06

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Carotenoid biosynthesis	711.45	1.47	0.32	4.58	4.57E-06	1.00E-05
	Carbohydrate digestion and absorption	8435.89	1.24	0.27	4.54	5.66E-06	1.22E-05
	Pores ion channels	127239.95	0.48	0.11	4.4	1.09E-05	2.29E-05
	Bile secretion	84.17	1.69	0.4	4.2	2.70E-05	5.65E-05
	Terpenoid backbone biosynthesis	220257.82	0.15	0.04	4.15	3.32E-05	6.89E-05
	Riboflavin metabolism	101662.66	0.19	0.05	4.05	5.21E-05	1.07E-04
	Pantothenate and CoA biosynthesis	217155.25	0.1	0.03	4.02	5.75E-05	1.18E-04
	Toluene degradation	36589.09	0.42	0.1	3.99	6.65E-05	1.35E-04
	RNA degradation	169545.41	0.09	0.02	3.92	8.81E-05	1.78E-04
	Isoquinoline alkaloid biosynthesis	23171.72	0.51	0.13	3.86	1.12E-04	2.23E-04
	Ubiquinone and other terpenoid-quinone biosynthesis	83378.49	0.53	0.14	3.77	1.65E-04	3.25E-04
	Nucleotide excision repair	149267.74	0.16	0.04	3.63	2.87E-04	5.53E-04
	Fluorobenzoate degradation	558.5	1.14	0.32	3.57	3.56E-04	6.71E-04
	Peptidases	668222.13	0.1	0.03	3.54	3.94E-04	7.38E-04
	Phenylalanine metabolism	50940.21	0.28	0.08	3.51	4.54E-04	8.45E-04
	Biosynthesis of ansamycins	31564.83	0.22	0.07	3.4	6.83E-04	1.25E-03
	Chlorocyclohexane and chlorobenzene degradation	1891.63	0.73	0.22	3.29	9.91E-04	1.79E-03
	Retinol metabolism	9076.56	0.6	0.19	3.19	1.45E-03	2.58E-03
	Folate biosynthesis	154779.27	0.17	0.05	3.15	1.61E-03	2.84E-03
	Homologous recombination	359146.91	0.06	0.02	3.08	2.04E-03	3.53E-03
	D-Arginine and D-ornithine metabolism	717.52	0.93	0.31	3.04	2.33E-03	4.01E-03
	Chromosome	598893.43	0.07	0.02	3.01	2.61E-03	4.43E-03
	Sulfur metabolism	87081.22	0.13	0.04	2.95	3.15E-03	5.21E-03
	Energy metabolism	291380.81	0.17	0.06	2.74	6.11E-03	9.81E-03

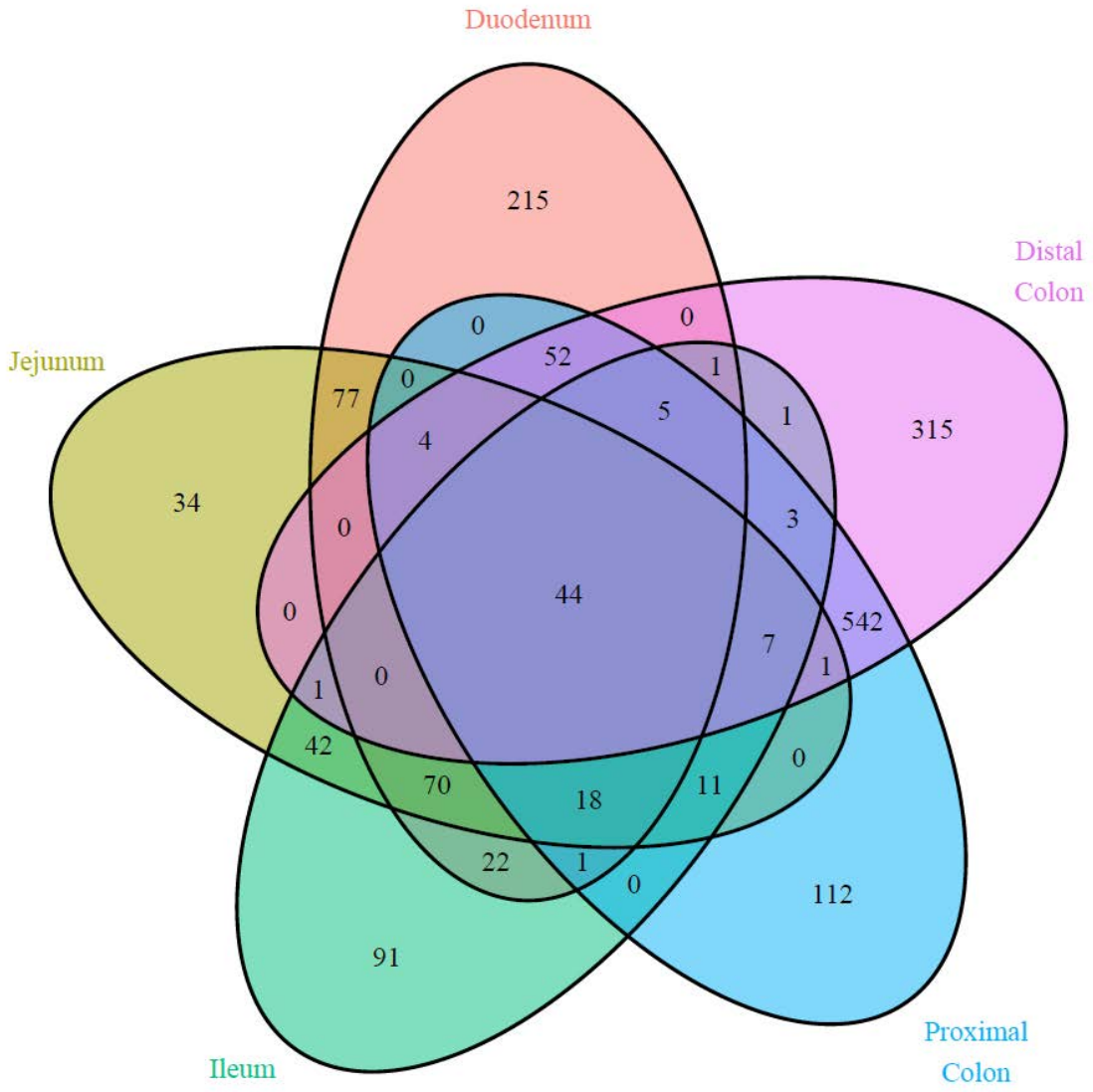
Proximal colon vs distal colon

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Lipopolysaccharide biosynthesis	130069.55	-0.28	0.03	-8.2	2.43E-16	3.19E-14
	Folate biosynthesis	137018.36	-0.1	0.01	-8.06	7.36E-16	6.45E-14
	Lipopolysaccharide biosynthesis proteins	156401.34	-0.24	0.03	-7.97	1.59E-15	1.05E-13
	Pores ion channels	120650.87	-0.17	0.02	-7.75	9.29E-15	4.07E-13
	1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation	389.58	-0.57	0.08	-7.46	8.69E-14	3.27E-12
	Metabolism of cofactors and vitamins	34028.28	-0.16	0.02	-7.19	6.32E-13	1.51E-11
	Riboflavin metabolism	89694.19	-0.14	0.02	-6.84	7.88E-12	1.48E-10
	Styrene degradation	4073.14	-0.4	0.06	-6.45	1.10E-10	1.38E-09
	Glutathione metabolism	67249.27	-0.16	0.03	-5.95	2.70E-09	2.73E-08
	Bladder cancer	459.54	-0.44	0.08	-5.86	4.77E-09	4.64E-08
	Prion diseases	609.8	-0.45	0.08	-5.77	8.10E-09	7.34E-08
	Various types of N-glycan biosynthesis	451.29	-0.43	0.08	-5.66	1.48E-08	1.26E-07
	Arachidonic acid metabolism	20238.29	-0.27	0.05	-5.47	4.52E-08	3.60E-07
	Isoquinoline alkaloid biosynthesis	22024.82	-0.19	0.03	-5.41	6.35E-08	4.91E-07
	Protein folding and associated processing	195108.49	-0.07	0.01	-5.08	3.85E-07	2.47E-06
	Other ion-coupled transporters	364490.09	-0.1	0.02	-5.01	5.31E-07	3.32E-06
	D-Arginine and D-ornithine metabolism	669.27	-0.38	0.08	-4.88	1.06E-06	6.04E-06
	Glycan biosynthesis and metabolism	13699.93	-0.2	0.04	-4.84	1.29E-06	7.21E-06
	Prenyltransferases	118436.11	-0.07	0.01	-4.59	4.53E-06	2.43E-05
	Biosynthesis of vancomycin group antibiotics	23245.58	-0.1	0.02	-4.48	7.30E-06	3.84E-05
	Vibrio cholerae pathogenic cycle	26957.13	-0.13	0.03	-4.47	7.81E-06	4.03E-05
	Translation proteins	294844.21	-0.03	0.01	-4.44	9.13E-06	4.53E-05
	Cardiac muscle contraction	1762.03	-0.33	0.08	-4.41	1.02E-05	4.91E-05
	Purine metabolism	726735.34	-0.05	0.01	-4.41	1.03E-05	4.91E-05
	Toluene degradation	34617.04	-0.12	0.03	-4.37	1.26E-05	5.93E-05
	Glycosyltransferases	126349.07	-0.08	0.02	-4.36	1.30E-05	5.98E-05
	Ubiquinone and other terpenoid-quinone biosynthesis	80418	-0.15	0.03	-4.32	1.59E-05	7.21E-05
	Membrane and intracellular structural molecules	200577.34	-0.09	0.02	-4.26	2.07E-05	8.91E-05
	Parkinsons disease	1876.94	-0.31	0.08	-4.16	3.24E-05	1.32E-04
	Chagas disease (American trypanosomiasis)	19.15	-0.19	0.05	-4.12	3.77E-05	1.46E-04
	Cellular antigens	21590.06	-0.14	0.04	-4.08	4.56E-05	1.67E-04
	Carbon fixation in photosynthetic organisms	201735.01	-0.04	0.01	-4.08	4.54E-05	1.67E-04
	Porphyrin and chlorophyll metabolism	236775.11	-0.11	0.03	-3.98	6.95E-05	2.50E-04
	Amino acid related enzymes	485598.68	-0.03	0.01	-3.88	1.05E-04	3.68E-04

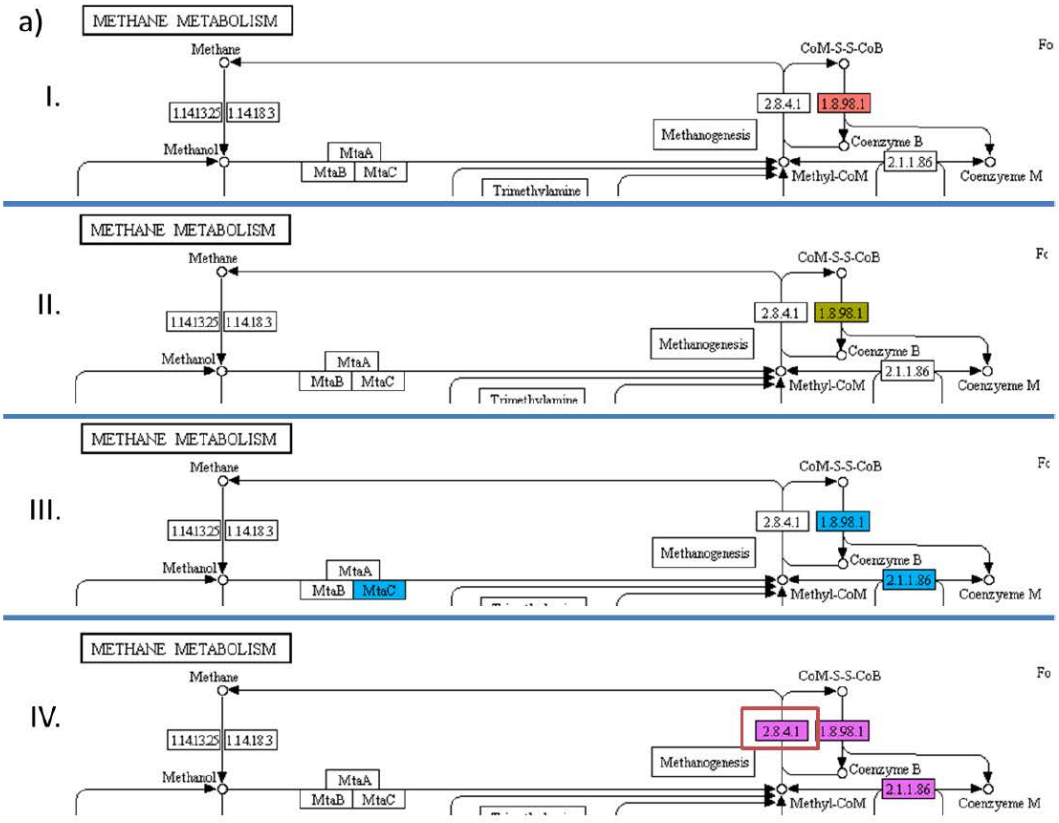
	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	alpha-Linolenic acid metabolism	1322.95	-0.3	0.08	-3.83	1.26E-04	4.30E-04
	Chaperones and folding catalysts	344621.37	-0.05	0.01	-3.79	1.49E-04	5.01E-04
	Protein digestion and absorption	13068.81	-0.15	0.04	-3.73	1.89E-04	6.21E-04
	Tropane, piperidine and pyridine alkaloid biosynthesis	37305.26	-0.07	0.02	-3.64	2.78E-04	8.50E-04
	Polyketide sugar unit biosynthesis	66802.58	-0.06	0.02	-3.48	5.02E-04	1.48E-03
	Terpenoid backbone biosynthesis	198223.99	-0.04	0.01	-3.37	7.65E-04	2.16E-03
	Selenocompound metabolism	113363.04	-0.04	0.01	-3.35	8.19E-04	2.27E-03
	Ribosome	820876.83	-0.03	0.01	-3.34	8.32E-04	2.28E-03
	Photosynthesis proteins	135219.79	-0.07	0.02	-3.33	8.57E-04	2.32E-03
	Type II diabetes mellitus	17449.45	-0.1	0.03	-3.32	9.04E-04	2.40E-03
	Zeatin biosynthesis	22133.68	-0.07	0.02	-3.32	9.00E-04	2.40E-03
	Cell cycle - Caulobacter	175302.74	-0.04	0.01	-3.31	9.47E-04	2.49E-03
	Huntingtons disease	7275.08	-0.23	0.07	-3.28	1.03E-03	2.67E-03
	Alzheimers disease	17838.02	-0.1	0.03	-3.28	1.05E-03	2.69E-03
	Photosynthesis	134276.25	-0.07	0.02	-3.28	1.05E-03	2.69E-03
	Mineral absorption	4909.22	-0.18	0.06	-3.26	1.12E-03	2.83E-03
	Carbon fixation pathways in prokaryotes	324682.92	-0.03	0.01	-3.22	1.28E-03	3.21E-03
	One carbon pool by folate	223923.24	-0.04	0.01	-3.18	1.47E-03	3.61E-03
	RNA polymerase	52852.2	-0.05	0.01	-3.05	2.28E-03	5.03E-03
	Chromosome	525062.14	-0.04	0.01	-2.94	3.24E-03	6.98E-03
	Carbohydrate digestion and absorption	10003.82	-0.14	0.05	-2.94	3.28E-03	7.02E-03
	Biosynthesis of siderophore group nonribosomal peptides	6367.76	-0.13	0.05	-2.88	3.94E-03	8.29E-03
	beta-Alanine metabolism	63276.58	-0.07	0.02	-2.88	3.92E-03	8.29E-03
	Bacterial secretion system	182726.33	-0.04	0.02	-2.85	4.40E-03	9.19E-03
	Pentose and glucuronate interconversions	123089.36	0.15	0.02	8.28	1.26E-16	3.19E-14
	Amoebiasis	1210.1	0.62	0.08	7.9	2.84E-15	1.49E-13
	Butirosin and neomycin biosynthesis	21773.81	0.16	0.02	7.35	1.95E-13	6.41E-12
	Protein kinases	68095.37	0.16	0.02	7.32	2.56E-13	7.47E-12
	Flavonoid biosynthesis	741.54	0.55	0.08	7.21	5.53E-13	1.46E-11
	Carbohydrate metabolism	37365.83	0.21	0.03	7.18	7.05E-13	1.54E-11
	Methane metabolism	332793.23	0.11	0.02	6.9	5.35E-12	1.08E-10
	Bile secretion	292.96	0.46	0.07	6.69	2.22E-11	3.90E-10
	Sphingolipid metabolism	55497.86	0.19	0.03	6.62	3.50E-11	5.75E-10
	Two-component system	371945.14	0.13	0.02	6.55	5.88E-11	8.13E-10
	Nitrotoluene degradation	17214.14	0.26	0.04	6.55	5.74E-11	8.13E-10
	Basal transcription factors	457.16	0.47	0.07	6.55	5.75E-11	8.13E-10
	Germination	8074.16	0.4	0.06	6.45	1.10E-10	1.38E-09

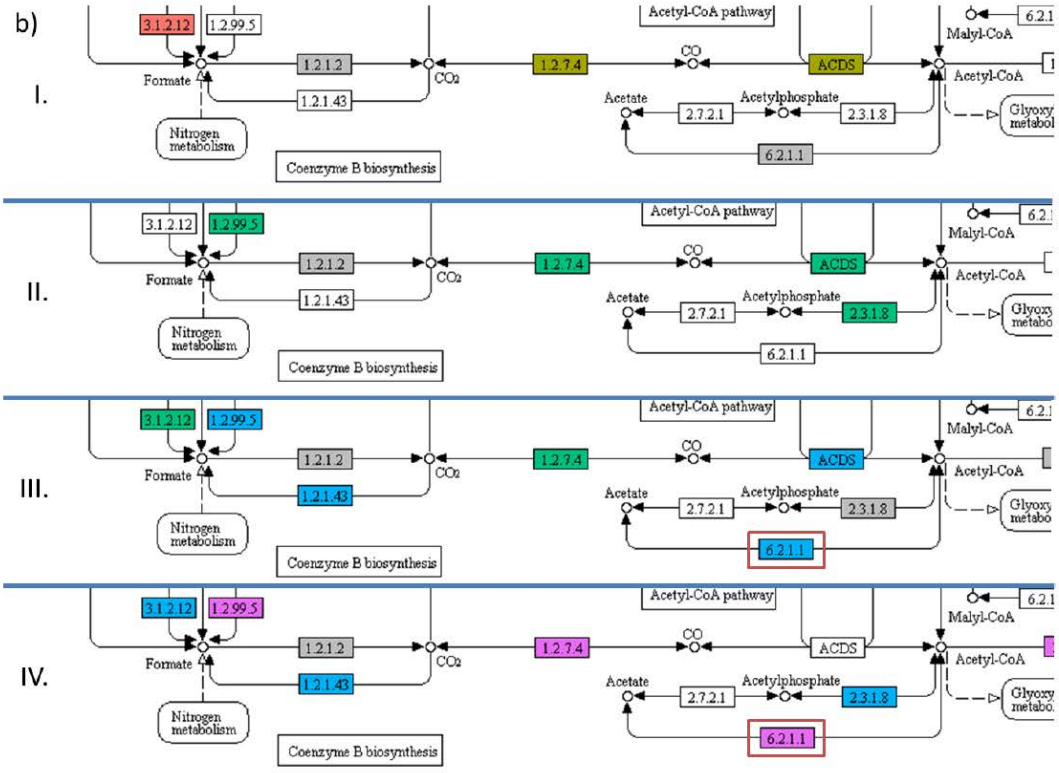
	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Steroid biosynthesis	606.8	0.33	0.05	6.31	2.76E-10	3.30E-09
	Proximal tubule bicarbonate reclamation	2926.52	0.46	0.07	6.26	3.90E-10	4.46E-09
	Lysosome	31906.67	0.22	0.04	6.04	1.53E-09	1.68E-08
	Others	228109.17	0.09	0.02	6.02	1.70E-09	1.79E-08
	Pyruvate metabolism	259924.12	0.07	0.01	5.85	5.00E-09	4.70E-08
	Primary bile acid biosynthesis	5666.76	0.36	0.06	5.75	8.75E-09	7.67E-08
	Sporulation	143944.78	0.24	0.04	5.49	4.11E-08	3.38E-07
	Systemic lupus erythematosus	95.37	0.24	0.05	5.28	1.26E-07	9.49E-07
	Secondary bile acid biosynthesis	5551.63	0.33	0.06	5.24	1.57E-07	1.15E-06
	Chloroalkane and chloroalkene degradation	34081.11	0.24	0.05	5.23	1.68E-07	1.20E-06
	PPAR signaling pathway	32644.55	0.1	0.02	5.23	1.74E-07	1.20E-06
	Bacterial chemotaxis	154250.37	0.2	0.04	5.12	3.02E-07	2.04E-06
	RNA transport	35034.26	0.12	0.02	5.1	3.41E-07	2.24E-06
	Other glycan degradation	76100.47	0.15	0.03	5	5.69E-07	3.48E-06
	Signal transduction mechanisms	119595.21	0.08	0.02	4.98	6.39E-07	3.82E-06
	Circadian rhythm - plant	93.56	0.21	0.04	4.95	7.35E-07	4.29E-06
	Phenylpropanoid biosynthesis	43083.39	0.24	0.05	4.75	2.07E-06	1.14E-05
	Transporters	1511492.19	0.09	0.02	4.45	8.54E-06	4.32E-05
	Glycosphingolipid biosynthesis - globo series	36281.7	0.12	0.03	4.29	1.79E-05	7.99E-05
	Xylene degradation	7431.94	0.31	0.07	4.28	1.89E-05	8.28E-05
	Valine, leucine and isoleucine biosynthesis	193609.46	0.07	0.02	4.21	2.60E-05	1.10E-04
	Glycolysis / Gluconeogenesis	307259.55	0.05	0.01	4.19	2.75E-05	1.15E-04
	Energy metabolism	270746.52	0.05	0.01	4.15	3.27E-05	1.32E-04
	Phosphonate and phosphinate metabolism	12143.16	0.17	0.04	4.15	3.39E-05	1.35E-04
	Starch and sucrose metabolism	287025.76	0.09	0.02	4.12	3.74E-05	1.46E-04
	Transcription machinery	294709.78	0.06	0.01	4.11	3.96E-05	1.51E-04
	Arginine and proline metabolism	317865.89	0.04	0.01	4.1	4.15E-05	1.56E-04
	Glycosaminoglycan degradation	23083.13	0.14	0.03	3.93	8.42E-05	2.99E-04
	Linoleic acid metabolism	13615.42	0.12	0.03	3.86	1.14E-04	3.93E-04
	Meiosis - yeast	2549.65	0.3	0.08	3.77	1.63E-04	5.42E-04
	Galactose metabolism	191618.98	0.09	0.02	3.72	2.00E-04	6.50E-04
	Naphthalene degradation	32381.88	0.08	0.02	3.68	2.37E-04	7.59E-04
	Phenylalanine metabolism	49699.13	0.08	0.02	3.66	2.49E-04	7.88E-04
	Adipocytokine signaling pathway	22838.59	0.09	0.03	3.65	2.61E-04	8.19E-04
	Glycerolipid metabolism	94293.18	0.07	0.02	3.64	2.72E-04	8.42E-04
	Lipid biosynthesis proteins	174917.22	0.06	0.02	3.56	3.69E-04	1.11E-03
	Histidine metabolism	166509.2	0.06	0.02	3.49	4.74E-04	1.42E-03

	KEGG PATHWAYS	base Mean	log2 Fold Change	lfcSE	stat	*p*-value	padj
	Type I diabetes mellitus	16620.1	0.06	0.02	3.46	5.42E-04	1.58E-03
	Inositol phosphate metabolism	22491.18	0.09	0.03	3.42	6.22E-04	1.80E-03
	Transcription related proteins	892.13	0.25	0.07	3.41	6.61E-04	1.89E-03
	Lipoic acid metabolism	6461.84	0.24	0.07	3.36	7.75E-04	2.17E-03
	Penicillin and cephalosporin biosynthesis	4977.03	0.19	0.06	3.21	1.31E-03	3.25E-03
	Bacterial motility proteins	304129.64	0.18	0.06	3.15	1.64E-03	4.00E-03
	Cyanoamino acid metabolism	87546.05	0.08	0.03	3.14	1.67E-03	4.02E-03
	Glycerophospholipid metabolism	149777.35	0.03	0.01	3.11	1.86E-03	4.45E-03
	Transcription factors	387127.44	0.06	0.02	3.11	1.88E-03	4.45E-03
	Cell cycle	24.45	0.11	0.03	3.09	2.02E-03	4.55E-03
	Hepatitis C	24.45	0.11	0.03	3.09	2.02E-03	4.55E-03
	Measles	24.45	0.11	0.03	3.09	2.02E-03	4.55E-03
	Phagosome	24.45	0.11	0.03	3.09	2.02E-03	4.55E-03
	Vibrio cholerae infection	24.45	0.11	0.03	3.09	2.02E-03	4.55E-03
	mTOR signaling pathway	24.45	0.11	0.03	3.09	2.02E-03	4.55E-03
	mRNA surveillance pathway	48.91	0.1	0.03	3.07	2.17E-03	4.83E-03
	Dioxin degradation	8801.3	0.22	0.07	3.01	2.64E-03	5.78E-03
	Peroxisome	50642.55	0.07	0.02	2.95	3.14E-03	6.83E-03



Supplementary Figure S1. Five-part Venn diagram performed for the OTUs shared among sections when combining the datasets from all subjects: duodenum (red), jejunum (yellow), ileum (green), proximal colon (blue), and distal colon (purple).


a)
I.
II.
III.
IV.
METHANE METABOLISM
Methane
Methanol
1.14.13.25
1.14.18.3
MtaA
MtaB
MtaC
Trimethylamine
CoM-S-S-CoB
2.8.4.1
1.8.98.1
Methanogenesis
Coenzyme B
2.1.1.86
Methyl-CoM
Coenzyme M


b)
I.
II.
III.
IV.
3.1.2.12
1.2.99.5
Formate
1.2.1.2
CO2
1.2.1.43
Nitrogen metabolism
Coenzyme B biosynthesis
1.2.7.4
CO
Acetyl-CoA pathway
ACDS
Malyl-CoA
Acetyl-CoA
Acetate
2.7.2.1
Acetylphosphate
2.3.1.8
6.2.1.1

Supplementary Figure S2. DESeq2[27] results below a padj≤0.01 cut-off for the four comparisons between each pair of consecutive sections of the KEGG[26] orthologies (KOs) predicted by PICRUSt[25] represented over the KEGG[26] methane metabolism pathway (map00680): I. duodenum vs jejunum; II. jejunum vs ileum; III. ileum vs proximal colon; IV. proximal colon vs distal colon. The colour shows which KO was more abundant in that section when comparing the two sections: white, non-significant; red, duodenum; yellow, jejunum; green, ileum; blue, proximal colon and purple, distal colon. Ambiguous KOs were coloured grey. For clarity of presentation, the methane metabolism pathway was divided into two parts: a) The red rectangle shows how the production of methane was more abundant in the distal colon than in the rest of the comparisons. b) The two red rectangles represent how the production of acetate was more abundant in the proximal colon than in the ileum and more abundant in the distal colon than in the proximal colon.

7.4 Supplementary material Paper IV: 'Association between the pig genome and its gut microbiota composition'

Supplementary Information S1. Full description of the 16S rRNA gene amplification and sequencing.

The following primers were used to amplify the V3-V4 region of the 16S rRNA gene as stated by Klindworth *et al*. (2013) and the *16S Metagenomic Sequencing Library Preparation* guide (Illumina, San Diego, CA, USA): Forward = 5' TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG 3' and Reverse = 5' GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC 3'. Each one of the 288 PCR reactions were carried out individually in a total volume of 25 µL utilizing 12.5 ng of microbial DNA, 12.5 µL of 2× KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) and 5 µL of each primer (1 µM) with the following protocol: 95ºC for 3 min, 25 cycles of three steps (95ºC for 30 s, 55ºC for 30 s and 72ºC for 30 s) and 72ºC for 5 min. The amplicon expected size (~550 bp) was verified via agarose gel electrophoresis. Then, the PCR product clean-up was performed with AMPure XP beads (Beckman Coulter, Beverly, MA, USA). After this step, the dual indices were attached with the Nextera XT Index Kit and another PCR clean-up round was performed with AMPure XP beads afterwards.

References

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013; **41**: 1–11.

Supplementary Table S1. Means for the relative abundance of the 18 phyla found in rectal contents of 285 pigs. %_Presence indicates the percentage of the pigs where these phyla were found.

Phylum	Mean	SD	%_Presence
p__Firmicutes	45.362674	5.3033424	100
p__Bacteroidetes	37.473308	4.5014274	100
p__Spirochaetes	6.8627622	1.9507945	100
p__Verrucomicrobia	2.8283591	1.8884845	100
p__Proteobacteria	2.5041132	0.9661273	100
p__Planctomycetes	1.1558683	0.802777	100
p__Fibrobacteres	1.0716888	0.7828378	100
p__Cyanobacteria	0.8856286	0.6531136	100
p__Tenericutes	0.7540225	0.3834549	100
p__Euryarchaeota	0.2982062	0.2533546	100
p__Fusobacteria	0.1585777	0.7286231	57.894737
p__Lentisphaerae	0.1420824	0.0718945	100
p__Elusimicrobia	0.1345645	0.1508312	99.649123
p__TM7	0.1257641	0.1002946	99.298246
p__Actinobacteria	0.1220437	0.0687014	100
p__Synergistetes	0.0575701	0.0470184	100
p__Deferribacteres	0.037369	0.0638809	88.77193
p__WPS-2	0.0253982	0.0348928	83.157895

Supplementary Table S2. Means for the relative abundance of the 101 genera found in rectal contents of 285 pigs. %_Presence indicates the percentage of the pigs where these genera were found.

Genus	Mean	SD	%_Presence
f__Ruminococcaceae.g__unsp	14.1274536	3.34371269	100.00
o__Bacteroidales.f__unsp.g__unsp	11.0394799	2.44254691	100.00
o__Clostridiales.f__unsp.g__unsp	7.6170881	1.49831863	100.00
g__Prevotella	7.02958386	2.32974885	100.00
g__Treponema	6.28989487	1.90171129	100.00
f__S24-7.g__unsp	4.3303154	1.29156961	100.00
f__Clostridiaceae.g__unsp	4.30869944	1.89437357	100.00
f__Lachnospiraceae.g__unsp	3.88667642	0.93522396	100.00
g__[Prevotella]	3.44352368	1.88210544	100.00
g__Lactobacillus	2.95196875	2.30782056	100.00
g__CF231	2.31181537	0.87339284	100.00
f__p-2534-18B5.g__unsp	1.92319023	1.30113955	100.00
f__RF16.g__unsp	1.86170702	0.86249023	100.00
f__RFP12.g__unsp	1.63922578	0.60446035	100.00
g__Oscillospira	1.57028019	0.41973981	100.00
g__Parabacteroides	1.51004471	0.6573709	100.00
f__BS11.g__unsp	1.29894812	1.27410596	98.25
g__Phascolarctobacterium	1.29608652	0.43017031	100.00
g__Bacteroides	1.20851987	0.94879448	100.00
g__SMB53	1.19404527	0.49895698	100.00
f__Christensenellaceae.g__unsp	1.16531328	1.41990441	100.00
f__Pirellulaceae.g__unsp	1.15586828	0.80277704	100.00
g__Ruminococcus	1.15169234	0.43753361	100.00
g__Fibrobacter	1.07168876	0.78283781	100.00
o__YS2.f__unsp.g__unsp	0.88562859	0.65311362	100.00
g__Coprococcus	0.87058399	0.34365739	100.00
g__Streptococcus	0.86769252	1.28542385	99.65
g__RFN20	0.65590069	0.3087897	100.00
g__Akkermansia	0.57297394	0.89875957	97.54
g__Sphaerochaeta	0.57286733	0.28000211	100.00
f__[Cerasicoccaceae].g__unsp	0.56965639	1.16582321	89.47
g__YRC22	0.51735383	0.42371428	100.00
g__Roseburia	0.48439181	0.22128081	100.00
g__Campylobacter	0.476466	0.46340464	100.00
f__[Paraprevotellaceae].g__unsp	0.46428294	0.24503345	99.65
g__Paludibacter	0.46396199	0.31203002	100.00

g__Turicibacter	0.44188495	0.24399812	100.00
g__p-75-a5	0.43076753	0.25550589	100.00
o__GMD14H09.f__unsp.g__unsp	0.42572947	0.54064667	98.60
g__Succinivibrio	0.38725412	0.39307622	100.00
g__Clostridium	0.37141741	0.18629842	100.00
g__Dorea	0.35639157	0.15886797	100.00
f__Erysipelotrichaceae.g__unsp	0.33138423	0.14091958	100.00
f__[Mogibacteriaceae].g__unsp	0.32848538	0.10231591	100.00
g__Anaeroplasma	0.31083365	0.20867746	99.65
o__RF32.f__unsp.g__unsp	0.28148621	0.2151074	99.65
g__Anaerovibrio	0.25487801	0.17865508	100.00
f__Mycoplasmataceae.g__unsp	0.19056224	0.2275562	97.19
g__Methanobrevibacter	0.17812547	0.21506095	91.23
g__Sutterella	0.17261243	0.11939029	100.00
g__Desulfovibrio	0.17017957	0.09267362	100.00
g__Fusobacterium	0.15857766	0.72862309	57.89
g__Flexispira	0.13160544	0.16419721	99.65
c__Alphaproteobacteria.o__unsp.f__unsp.g__unsp	0.12923124	0.11600164	99.65
f__F16.g__unsp	0.12576414	0.10029461	99.30
o__RF39.f__unsp.g__unsp	0.12297434	0.07904396	100.00
g__vadinCA11	0.12008074	0.07899679	100.00
g__rc4-4	0.10356609	0.05614116	100.00
f__Elusimicrobiaceae.g__unsp	0.09362324	0.10405143	98.60
g__Lachnospira	0.09147395	0.09382415	99.30
f__Desulfovibrionaceae.g__unsp	0.08714971	0.03830128	100.00
g__Bulleidia	0.08675445	0.08112611	98.60
g__L7A_E11	0.0814239	0.0525118	100.00
o__Tremblayales.f__unsp.g__unsp	0.07995681	0.14059091	89.82
f__R4-45B.g__unsp	0.07790942	0.05056144	99.65
f__Enterobacteriaceae.g__unsp	0.07577719	0.29158587	74.04
g__Blautia	0.07058708	0.06996154	99.65
f__Coriobacteriaceae.g__unsp	0.06572288	0.03489308	100.00
f__Victivallaceae.g__unsp	0.06417299	0.04151603	100.00
o__ML615J-28.f__unsp.g__unsp	0.06356403	0.11105417	95.79
f__Rikenellaceae.g__unsp	0.06044513	0.05814588	97.89
g__Pyramidobacter	0.04848042	0.02398795	100.00
f__WCHB1-25.g__unsp	0.04650295	0.0521698	93.68
g__Elusimicrobium	0.04094123	0.08535976	75.79
f__Anaeroplasmataceae.g__unsp	0.03811668	0.02795448	99.30
g__Mucispirillum	0.03736902	0.06388086	88.77

g__Caloramator	0.03705362	0.2854699	15.44
g__Ureibacillus	0.03596477	0.40583577	21.40
g__Epulopiscium	0.03366484	0.02572925	99.65
g__Helicobacter	0.03015033	0.03705317	95.09
g__Anaerovorax	0.02997118	0.02575723	98.95
g__Bifidobacterium	0.02671872	0.0484876	70.53
o__Rickettsiales.f__unsp.g__unsp	0.02600314	0.05167286	86.32
p__WPS-2.c__unsp.o__unsp.f__unsp.g__unsp	0.02539816	0.03489275	83.16
g__Adlercreutzia	0.02406793	0.01694647	98.95
g__Ruminobacter	0.02243522	0.02691804	92.28
g__[Eubacterium]	0.01997109	0.04118542	65.26
g__Selenomonas	0.01918839	0.03567601	74.39
g__Tepidimicrobium	0.01771429	0.1544598	11.93
g__Anoxybacillus	0.01613449	0.1247667	25.26
g__Mycoplasma	0.01426743	0.05840599	52.28
o__Acholeplasmatales.f__unsp.g__unsp	0.01370418	0.01394597	89.82
f__Peptostreptococcaceae.g__unsp	0.01362976	0.01782999	95.09
g__Pseudobutyrivibrio	0.01358921	0.03759048	60.00
g__Peptococcus	0.01265677	0.01488801	68.77
g__Paraprevotella	0.01013555	0.01283101	80.00
f__Dethiosulfovibrionaceae.g__unsp	0.00908964	0.04024258	32.28
g__Bacillus	0.00866719	0.01762395	50.88
g__Yersinia	0.00807628	0.04383139	23.86
g__[Ruminococcus]	0.00758089	0.01259968	69.12
g__Slackia	0.00553419	0.0047794	81.75

Supplementary Table S3. Description of the 52 significant genera-associated SNPs and their predicted consequences with the Variant Effect Predictor tool (McLaren *et al.*, 2010) (Ensembl release 92). MAF indicates the minor allele frequency.

Region	SNP	Chr.	Position	MAF	*p*-value	FDR	Genus
A1	rs328868175	3	1,885,041	0.105	1.27E-05	7.55E-02	*Akkermansia*
A1	rs81335357	3	2,026,411	0.14	4.04E-06	4.58E-02	*Akkermansia*
A1	rs81246645	3	2,037,807	0.14	4.04E-06	4.58E-02	*Akkermansia*
A2	rs81390429	6	102,464,736	0.458	6.06E-06	4.58E-02	*Akkermansia*
A3	rs325604118	7	113,612,765	0.091	1.33E-05	7.55E-02	*Akkermansia*
A4	rs81410866	9	48,531,463	0.089	5.25E-06	4.58E-02	*Akkermansia*
A4	rs81410881	9	48,574,214	0.089	5.25E-06	4.58E-02	*Akkermansia*
A5	rs80982646	15	100,099,175	0.067	3.04E-06	4.58E-02	*Akkermansia*
B1	rs319005051	4	120,910,431	0.265	5.72E-06	6.48E-02	*CF231*
B2	rs344470822	23	113,482,584	0.243	4.72E-06	6.48E-02	*CF231*
B2	rs345545405	23	113,500,362	0.24	5.06E-06	6.48E-02	*CF231*
B2	rs329229283	23	113,504,919	0.242	4.10E-06	6.48E-02	*CF231*
C1	rs81223434	9	66,328,528	0.153	1.25E-06	5.68E-02	*Phascolarctobacterium*
D1	rs326174858	8	4,812,082	0.198	7.79E-07	3.53E-02	*Prevotella*
E1	rs81421752	10	19,512,266	0.298	3.28E-07	4.95E-03	*SMB53*
E1	rs344136854	10	19,535,825	0.216	4.92E-09	1.68E-04	*SMB53*
E1	rs81333406	10	20,444,762	0.196	7.43E-09	1.68E-04	*SMB53*
E1	rs332973121	10	20,906,151	0.233	7.55E-06	2.01E-02	*SMB53*
E1	rs81287953	10	21,054,756	0.219	1.65E-05	4.15E-02	*SMB53*
E2	rs341165563	10	54,894,603	0.204	3.15E-05	7.14E-02	*SMB53*
E3	rs81430839	11	29,219,531	0.239	3.56E-06	1.47E-02	*SMB53*
E3	rs81253096	11	30,233,589	0.251	2.06E-06	1.47E-02	*SMB53*
E3	rs80835110	11	30,525,947	0.246	1.37E-06	1.47E-02	*SMB53*
E3	rs80935884	11	30,624,084	0.266	5.59E-06	1.69E-02	*SMB53*
E3	rs80846807	11	30,966,247	0.265	6.08E-06	1.72E-02	*SMB53*
E3	rs80932731	11	31,222,268	0.239	3.56E-06	1.47E-02	*SMB53*
E3	rs326976142	11	31,296,557	0.239	3.56E-06	1.47E-02	*SMB53*
E3	rs80952844	11	31,389,565	0.239	3.56E-06	1.47E-02	*SMB53*
E3	rs81477799	11	31,932,828	0.239	3.56E-06	1.47E-02	*SMB53*
E3	rs81221998	11	32,016,286	0.263	5.41E-06	1.69E-02	*SMB53*
E3	rs81228368	11	32,256,752	0.263	5.41E-06	1.69E-02	*SMB53*
E3	rs81430905	11	32,274,205	0.263	5.41E-06	1.69E-02	*SMB53*
E3	rs81430907	11	32,313,997	0.267	2.84E-06	1.47E-02	*SMB53*
E3	rs81430928	11	32,504,246	0.261	1.82E-05	4.35E-02	*SMB53*

Region	SNP	Chr.	Position	MAF	*p*-value	FDR	Genus
F1	rs81277042	9	24,450,219	0.116	3.85E-05	9.86E-02	*Streptococcus*
F1	rs81407835	9	24,588,354	0.116	3.85E-05	9.86E-02	*Streptococcus*
F1	rs319168851	9	24,658,363	0.106	3.11E-05	9.41E-02	*Streptococcus*
F2	rs337860934	13	3,973,529	0.391	2.84E-05	9.41E-02	*Streptococcus*
F2	rs81310237	13	4,029,229	0.449	1.20E-05	9.41E-02	*Streptococcus*
F2	rs81338701	13	4,065,640	0.389	3.91E-05	9.86E-02	*Streptococcus*
F2	rs323657884	13	4,153,039	0.112	1.71E-05	9.41E-02	*Streptococcus*
F3	rs337448241	14	134,803,873	0.321	3.46E-06	9.41E-02	*Streptococcus*
F4	rs81344152	15	26,146,841	0.291	2.08E-05	9.41E-02	*Streptococcus*
F4	rs331341379	15	26,170,988	0.288	1.12E-05	9.41E-02	*Streptococcus*
F4	rs322132662	15	26,404,955	0.256	2.99E-05	9.41E-02	*Streptococcus*
F4	rs324157363	15	26,676,520	0.256	2.99E-05	9.41E-02	*Streptococcus*
F4	rs336637168	15	26,733,738	0.256	2.99E-05	9.41E-02	*Streptococcus*
F4	rs340766041	15	26,769,599	0.256	2.99E-05	9.41E-02	*Streptococcus*
F4	rs81452072	15	26,801,277	0.256	2.99E-05	9.41E-02	*Streptococcus*
F4	rs81235010	15	26,842,383	0.274	2.97E-05	9.41E-02	*Streptococcus*
F4	rs322734016	15	26,869,230	0.274	2.97E-05	9.41E-02	*Streptococcus*
F5	rs334064749	18	45,252,761	0.133	2.26E-05	9.41E-02	*Streptococcus*

Supplementary Table S3. continuation (*consequence prediction*)

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
A1	rs328868175	5_prime_UTR_variant; intron_variant; upstream_gene_variant	ENSSSCG00000007568	IQCE
A1	rs81335357	intron_variant	ENSSSCG00000032549	GNA12
A1	rs81246645	intron_variant	ENSSSCG00000032549	GNA12
A2	rs81390429	intergenic_variant	-	-
A3	rs325604118	3_prime_UTR_variant	ENSSSCG00000002446	-
A4	rs81410866	intron_variant	ENSSSCG00000015135	SORL1
A4	rs81410881	intron_variant	ENSSSCG00000015135	SORL1
A5	rs80982646	intron_variant	ENSSSCG00000016068	HECW2
B1	rs319005051	intergenic_variant	-	-
B2	rs344470822	intron_variant	ENSSSCG00000022331	FGF13
B2	rs345545405	intron_variant	ENSSSCG00000022331	FGF13
B2	rs329229283	intron_variant	ENSSSCG00000022331	FGF13
C1	rs81223434	intergenic_variant	-	-
D1	rs326174858	missense_variant	ENSSSCG00000032637	EVC

Region	SNP	Consequence	Ensembl_Gene_Id	Gene_Symbol
E1	rs81421752	intron_variant	ENSSSCG00000011176	SUSD4
E1	rs344136854	intron_variant	ENSSSCG00000011176	SUSD4
E1	rs81333406	intron_variant	ENSSSCG00000010900	DENND1B
E1	rs332973121	intergenic_variant	-	-
E1	rs81287953	intron_variant	ENSSSCG00000010904	NEK7
E2	rs341165563	intron_variant	ENSSSCG00000040615	MALRD1
E3	rs81430839	intergenic_variant	-	-
E3	rs81253096	intergenic_variant	-	-
E3	rs80835110	intron_variant	ENSSSCG00000009446	PCDH17
E3	rs80935884	intergenic_variant	-	-
E3	rs80846807	intergenic_variant	-	-
E3	rs80932731	intergenic_variant	-	-
E3	rs326976142	intergenic_variant	-	-
E3	rs80952844	intergenic_variant	-	-
E3	rs81477799	intergenic_variant	-	-
E3	rs81221998	intergenic_variant	-	-
E3	rs81228368	intergenic_variant	-	-
E3	rs81430905	intergenic_variant	-	-
E3	rs81430907	intergenic_variant	-	-
E3	rs81430928	intron_variant	ENSSSCG00000009448	DIAPH3
F1	rs81277042	intergenic_variant	-	-
F1	rs81407835	intergenic_variant	-	-
F1	rs319168851	intergenic_variant	-	-
F2	rs337860934	intron_variant	ENSSSCG00000029811	PLCL2
F2	rs81310237	intron_variant	ENSSSCG00000029811	PLCL2
F2	rs81338701	intron_variant	ENSSSCG00000029811	PLCL2
F2	rs323657884	intron_variant	ENSSSCG00000029811	PLCL2
F3	rs337448241	intergenic_variant	-	-
F4	rs81344152	intergenic_variant	-	-
F4	rs331341379	intergenic_variant	-	-
F4	rs322132662	intergenic_variant	-	-
F4	rs324157363	intergenic_variant	-	-
F4	rs336637168	intergenic_variant	-	-
F4	rs340766041	intergenic_variant	-	-
F4	rs81452072	intergenic_variant	-	-
F4	rs81235010	intergenic_variant	-	-
F4	rs322734016	intergenic_variant	-	-
F5	rs334064749	intergenic_variant	-	-

ACKNOWLEDGEMENTS