







SCIENTIFIC REPORTS



OPEN

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, Yulíaxis Ramayo-Caldas^{3,4}, Cristina Óvilo ⁵, Ana I. Fernández⁵, Maria Ballester ⁴ & Josep M. Folch^{1,2}

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.

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 nutrients¹. In humans, the genomes of these microorganisms (the so-called metagenome) contain more than 9 million unique genes². With the publication of the first reference gene catalogue of the porcine gut microbiome³, 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 humans⁴. Furthermore, both species share more non-redundant genes in their microbiota than humans do with other model organisms, such as the mouse³. Nevertheless, much of the work on the relationship between human obesity and the gut microbiota has been performed in mice⁵.

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 exponentially⁸⁻¹⁴. Interestingly, the pig gut microbiota composition has recently been related to average daily weight gain and to body weight^{10,13}, feed efficiency¹⁵, feed conversion and feed intake¹⁶. However, only a few studies have analysed

¹Plant and Animal Genomics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB Consortium, Bellaterra, Spain. ²Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona (UAB), Bellaterra, Spain. ³Génétique Animale et Biologie Intégrative (GABI), Institut National de la Recherche Agronomique (INRA), AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France. ⁴Departament de Genètica i Millora Animal, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Caldes de Montbui, Spain. ⁵Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain. Correspondence and requests for materials should be addressed to D.C.-P. (email: daniel.crespo@cragenomics.es)

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 White¹⁷, Laiwu¹⁸ and Gloucestershire Old Spot¹⁹ 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 acids²⁰. 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 pipeline²², and a total of 1,669 operational taxonomic units (OTUs) were obtained among the five sections. According to the Greengenes 13.8 database²³, 643 were new OTUs. In addition, the 1,669 OTUs were aggregated into 179 genera and 18 phyla (Fig. 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) (Fig. 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.

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 (Fig. 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 Fig. 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 (Fig. 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 (Fig. 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 (Fig. 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). 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).

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 (Fig. 4).

In the duodenum (Fig. 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

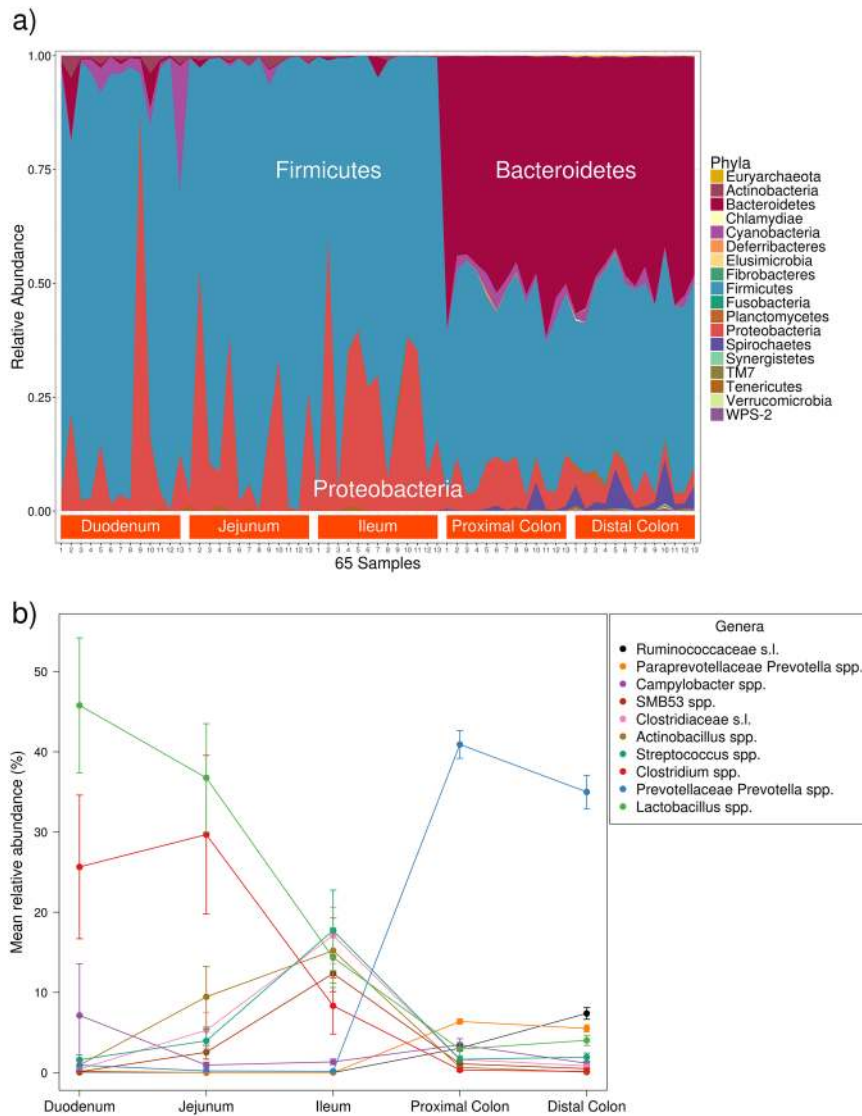


Figure 1. (a) Stacked area plot of the OTUs grouped by phyla for the 65 samples sorted by intestinal section. (b) Percentage evolution along the gut of the ten most abundant bacterial genera in the dataset. Segments represent the standard error.

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 (Fig. 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 (Fig. 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 (Fig. 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 (Fig. 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.

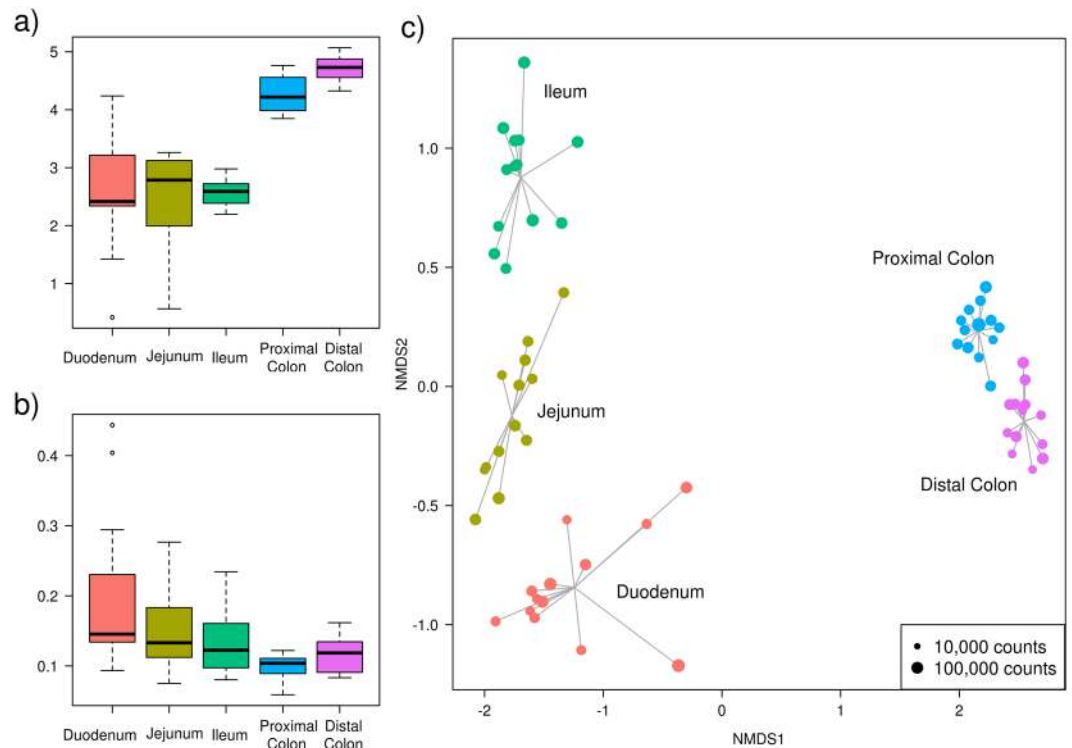


Figure 2. Descriptive plots made from the OTUs obtained in each sample. (a) Boxplot of the Shannon α -diversity for the 13 pigs in each intestinal section. (b) Boxplot of the Whittaker β -diversity for the 13 pigs in each intestinal section. (c) Non-metric multidimensional scaling (NMDS) plot based on 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.

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 metagenomeSeq²⁴. 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 Tables S1 and 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

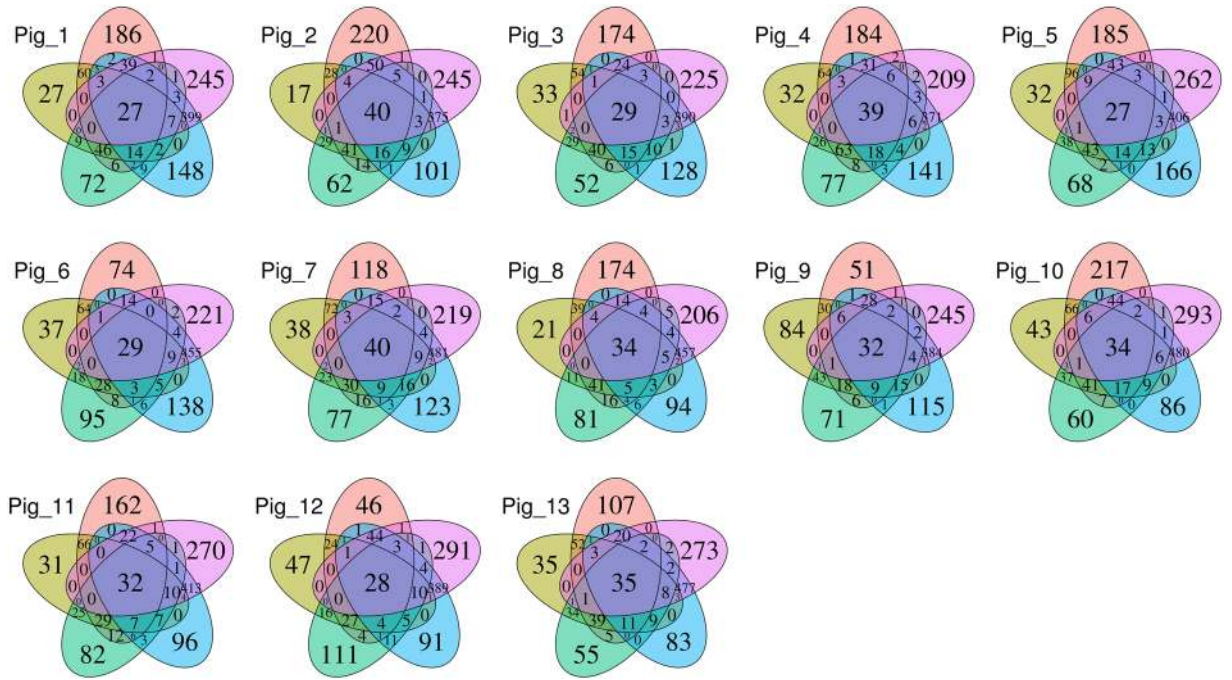


Figure 3. Five-part Venn diagram for each of the 13 subjects, showing the OTUs shared among the intestinal sections: duodenum (red), jejunum (yellow), ileum (green), proximal colon (blue), and distal colon (purple). The numbers in the diagrams represent how many OTUs were unique in the five intestinal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

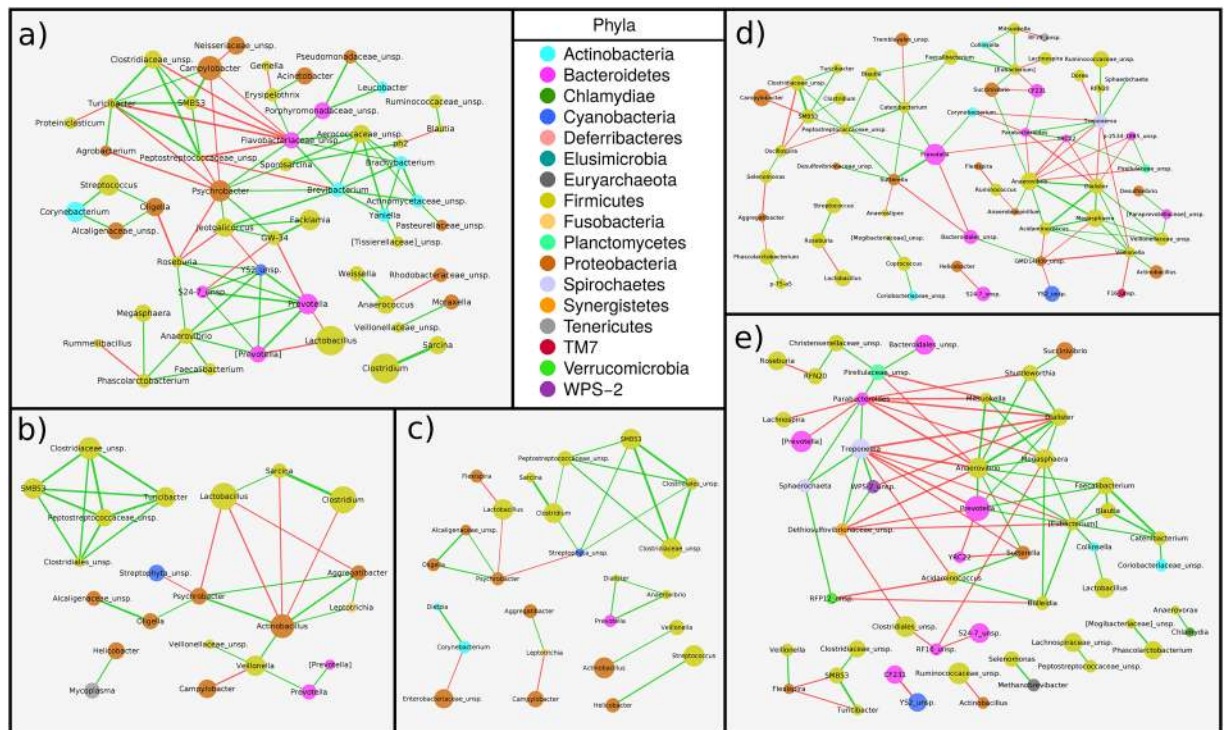


Figure 4. SparCC⁵⁵ PCIT⁵⁶ (SPCIT¹³) prokaryotic genus network for partial correlations with an absolute value above 0.65 between log-transformed genus abundances performed in the (a) duodenum, (b) jejunum, (c) ileum, (d) proximal colon, and (e) distal colon. The width of the edge represents the degree 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.

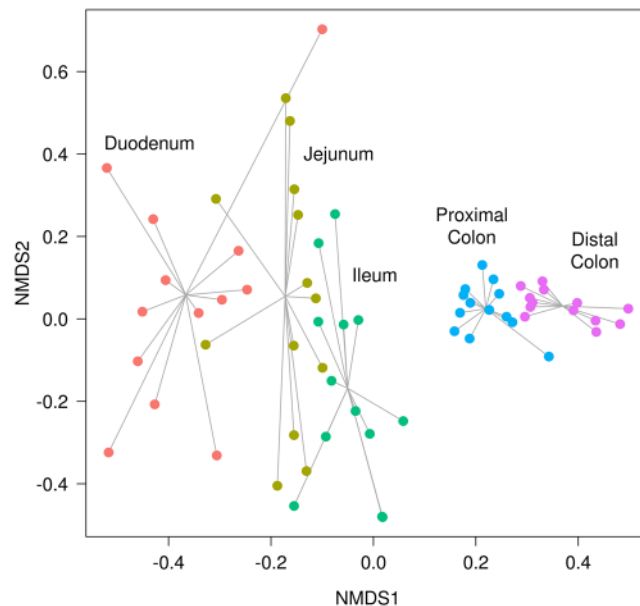


Figure 5. Non-metric multidimensional scaling (NMDS) plot based on Bray-Curtis dissimilarities for the metagenome (KEGG²⁶ orthologies (KOs) counts) predicted through PICRUSt²⁵ for the 65 samples from the 13 pigs in each of the 5 intestinal sections (represented by colours). In this plot, it can be seen 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.

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, PICRUSt²⁵ 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 KEGG²⁶ orthologies (KOs) table (Fig. 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 (Fig. 2b). Conversely, this NMDS plot (Fig. 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).

To gain a better understanding of the differentially abundant functions between the regions, the KOs were collapsed to the pathway level. Then, DESeq2²⁷ 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 glucuronate 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), 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 (Fig. 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 pigs¹⁹ from the Gloucestershire Old Spot breed. In addition, the β -diversity (Fig. 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 (Fig. 2c), where the points are closer in the hindgut, and in the smaller standard errors represented in the ten most abundant genera plot (Fig. 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 (Fig. 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 mucus²⁸. 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 unknown²⁸. 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 animals²⁹.

Consistent with the differences between intestinal sections and in accordance with other studies of the luminal^{17,18} and mucosal¹⁹ microbiota in pigs, the three most abundant phyla found along the digestive tract were Bacteroidetes, Firmicutes and Proteobacteria (Fig. 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^{30,31}, differentiating the microbiota found in these two regions¹⁹. 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 maize¹⁷ and 300-day-old Laiwu pigs fed with a maize-soybean diet¹⁸. In addition, it has been shown that different ages¹⁷, feeding³², and DNA extraction kits³³, 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 rate²⁰, 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 (Fig. 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). 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 intestine²⁸ and the different environmental conditions, such as oxygen concentrations^{30,31}, 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 sections²⁸. *Campylobacter* spp. are microaerophilic microorganisms, while most of the members of the Clostridiales order are obligate anaerobes³¹. In this context, the duodenum has a higher concentration of accessible oxygen³⁸, 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 anaerobes³⁹, 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 microbiota⁴¹. 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 Clostridiales⁴⁰. 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 maintenance⁴².

SCFA production is increased in the large intestine⁴³. 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 (Fig. 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 xylan⁴⁴. 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 SCFAs⁸ that can be absorbed by the host³⁵.

In our study, some Archaea related to methane metabolism were found to be differentially abundant in the hindgut, where methane is predominantly produced⁴³. Moreover, the methane concentration increases towards the end of the intestine and can be modified depending on dietary fibre content⁴⁵. Thus, methane metabolism was analysed in more detail (see Supplementary Fig. S2). 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' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG, and Reverse, 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATC. 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 2 \times KAPA HiFi HotStart ReadyMix (*Kapa Biosystems*, Inc.) and 5 μL of each primer (1 μM) with the following program: 95 $^{\circ}\text{C}$ for 3 minutes, 25 cycles of three steps (95 $^{\circ}\text{C}$ for 30 s, 55 $^{\circ}\text{C}$ for 30 s and 72 $^{\circ}\text{C}$ for 30 s) and 72 $^{\circ}\text{C}$ 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 pipeline²² (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 BLAST⁴⁹. Lastly, the final OTU dataset was obtained by filtering out singletons and OTUs representing less than 0.005% of the total counts in each section⁴⁷. 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 database²³ 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 package⁵⁰. In NMDS, the dissimilarity between pairs of samples was estimated with the Bray-Curtis distance⁵¹. The 5-part Venn diagrams were represented using the *draw.quintuple.venn* function of the VennDiagram R package⁵².

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 package⁵³. Then, the genera presence/absence analysis between each section and the next one was carried out with the metagenomeSeq R package²⁴

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) method⁵⁴.

Metagenome prediction and functional differences amongst gut sections. The metagenome KEGG²⁶ orthologies (KOs) of the 65 samples were predicted with the PICRUSt software²⁵, removing the OTUs that were not present in the Greengenes 13.5 database²³. Then, the Bray-Curtis distance⁵¹ was used to measure the dissimilarity between pairs of samples to create an NMDS plot of the KOs with the vegan package⁵⁰. 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 package²⁷ 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 method⁵⁴ 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) algorithm⁵⁶. The Cytoscape software⁵⁷ 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.

References

- 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).
- 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).
- Xiao, L. *et al.* A reference gene catalogue of the pig gut microbiome. *Nat. Microbiol.* **1**, 16161 (2016).
- Wang, M. & Donovan, S. M. Human microbiota-associated swine: current progress and future opportunities. *ILAR J.* **56**, 63–73 (2015).
- 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).
- Kuczynski, J. *et al.* Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* **13**, 47–58 (2011).
- 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).
- 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).
- 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).
- Mach, N. *et al.* Early-life establishment of the swine gut microbiome and impact on host phenotypes. *Environ. Microbiol. Rep.* **7**, 554–569 (2015).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Zhao, W. *et al.* The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments. *PLoS One* **10**, 1–13 (2015).
- 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).
- 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).
- Serra, X. *et al.* A comparison of carcass, meat quality and histochemical characteristics of Iberian (Guadyrbas 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. USA* **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. Fagarasan, S. *et al.* Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* **298**, 1424–7 (2002).
30. 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).
31. Albenberg, L. *et al.* Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. *Gastroenterology* **147**, 1055–1063.e8 (2014).
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. USA* **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).

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*.

Author Contributions

J.M.F. and A.I.F. conceived and designed the experiments; J.M.F. was the principal investigator of the project; this work is part of the PhD thesis of D.C.P., co-supervised by M.B. and J.M.F.; D.C.P., J.E., M.R., C.O. and J.M.F. collected samples; D.C.P., L.C.M. and M.B. performed the microbial DNA extraction; D.C.P., J.E. and Y.R.C. analysed the data; D.C.P., M.B. and J.M.F. wrote the paper. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30932-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018