environmental microbiology reports

Environmental Microbiology Reports (2015) **7**(3), 554–569 doi:10.1111/1758-2229.12285

Early-life establishment of the swine gut microbiome and impact on host phenotypes

Núria Mach,[1,2,3,4] Mustapha Berri,[5,6] Jordi Estellé,[3,4] Florence Levenez,[1,2] Gaëtan Lemonnier,[3,4] Catherine Denis,[3,4] Jean-Jacques Leplat,[3,4,7] Claire Chevaleyre,[5,6] Yvon Billon,[8] Joël Doré,[1,2] Claire Rogel-Gaillard[3,4†] and Patricia Lepage[1,2*†]

[1]*INRA, UMR1319 MICALIS, Jouy-en-Josas, France.*
[2]*AgroParisTech, UMR1319 MICALIS, Jouy-en-Josas, France.*
[3]*INRA, UMR 1313 Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France.*
[4]*AgroParisTech, UMR 1313 Génétique Animale et Biologie Intégrative, France.*
[5]*UMR1282 ISP, INRA, Nouzilly, France.*
[6]*UMR1282 ISP, Université de Tours, Tours, France.*
[7]*CEA, DSV-IRCM-LREG, Jouy-en-Josas, France.*
[8]*UE967 GEPA, INRA, Le Magneraud, France.*

Summary

Early bacterial colonization and succession within the gastrointestinal tract has been suggested to be crucial in the establishment of specific microbiota composition and the shaping of host phenotype. Here, the composition and dynamics of faecal microbiomes were studied for 31 healthy piglets across five age strata (days 14, 36, 48, 60 and 70 after birth) together with their mothers. Faecal microbiome composition was assessed by 16S rRNA gene 454-pyrosequencing. Bacteroidetes and Firmicutes were the predominant phyla present at each age. For all piglets, luminal secretory IgA concentration was measured at day 70, and body weight was recorded until day 70. The microbiota of suckling piglets was mainly represented by *Bacteroides*, *Oscillibacter, Escherichia/Shigella, Lactobacillus* and unclassified *Ruminococcaceae* genera. This pattern contrasted with that of *Acetivibrio, Dialister, Oribacterium, Succinivibrio* and *Prevotella* genera, which appeared increased after weaning. *Lactobacillus fermentum* might be vertically transferred via breast milk or faeces. The microbiota composition coevolved with their hosts towards two different clusters after weaning, primarily distinguished by unclassified *Ruminococcaceae* and *Prevotella* abundances. *Prevotella* was positively correlated with luminal secretory IgA concentrations, and body weight. Our study opens up new possibilities for health and feed efficiency manipulation via genetic selection and nutrition in the agricultural domain.

Received 11 December, 2014; accepted 22 February, 2015. *For correspondence. E-mail patricia.lepage@jouy.inra.fr; Tel. +33 134 652 344; Fax +33 134 652 521. †Equal contributors.

Introduction

The contribution of gastrointestinal tract microbiota to pig health and performance, including metabolism of nutrients, stimulation of immune response, protection from pathogens and stimulation of epithelium cell proliferation is becoming increasingly apparent (Katouli *et al.*, 1997a; Spreeuwenberg *et al.*, 2001; Konstantinov *et al.*, 2006; Lalles *et al.*, 2007; Thompson *et al.*, 2008; Mann *et al.*, 2014). In pigs, the microbial establishment begins at birth when the neonate is exposed to a wide variety of microorganisms, mainly provided by the mother during and after the passage through the birth canal and the surrounding environment (Katouli *et al.*, 1997b). Furthermore, since newborn piglets are in constant contact with the mother's faeces, skin and mucosal surfaces until weaning, it is likely that the establishment of their microbiota depends on that of the sow (Katouli *et al.*, 1997b; Thompson *et al.*, 2008). The age-related successional mechanisms and steps involved in the colonization and diversification of the microbiota in pigs are beginning to be understood (Inoue *et al.*, 2005; Konstantinov *et al.*, 2006; Thompson *et al.*, 2008; Kim *et al.*, 2011; Schmidt *et al.*, 2011; Buzoianu *et al.*, 2012; Looft *et al.*, 2012; Bearson *et al.*, 2013; Mann *et al.*, 2014; Schokker *et al.*, 2014).

However, little is known about how early-life establishment of the swine gut microbiome may contribute to the individual's performance, to the response to some medical and nutritional treatments and to the susceptibility to diseases. In humans, there is increasing evidence that perturbations of gut microbiota composition and functions may play an important role in the development of host metabolism and diseases (Clavel *et al.*, 2014). Current perception is that microbiota establishment is regulated by the metabolic niche (manly diet and antimicrobials), host genetic background, microbes–microbes interactions and host–microorganism interplay (Spor *et al.*, 2011; Schloss *et al.*, 2012; Bearson *et al.*, 2013). At a microbial-host regulatory level, a tight link between microbes and

IgA has been first described in axenic mice, in which colonization by a dense microbiota is mandatory for the production and secretion of luminal IgA (Benveniste *et al*., 1971). Moreover, dynamic secretion of IgA into the gut lumen might coat the bacteria helping maintain a tolerant, non-inflammatory host-microbial relationship (Macpherson, 2006; Levast *et al*., 2010; Sutherland and Fagarasan, 2012; Palm *et al*., 2014).

In this work, we present the first longitudinal study aimed at assessing how the pre- and post-weaning microbiota co-evolve with its host. We further investigated whether the lactation-adapted microbiota in healthy piglets presented stronger resemblance to their mother than to unrelated mothers. Lastly, we studied whether the microbiota composition and structure in early age can lead to the identification of microbial biomarkers associated with variations of luminal secretory IgA (sIgA) and growth traits.

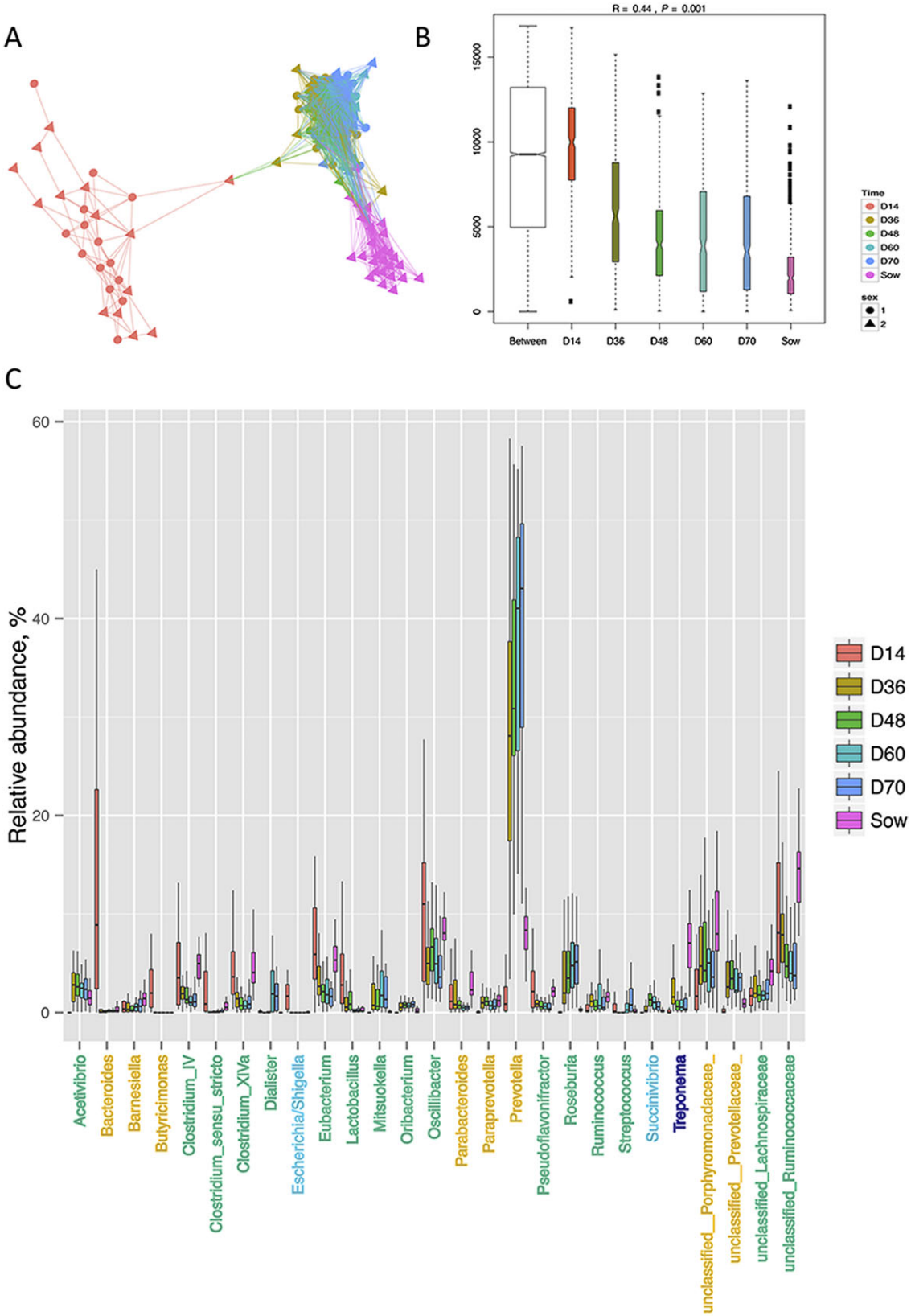
Results and discussion

Comparative analysis of microbiota composition and dynamics in piglets as a function of age, gender, pen and their mothers

Thirty-one large white piglets (15 females and 16 males) were studied together with their respective mothers ($n = 29$). One or two piglets from each litter was weaned at 28 days of age and randomly assigned to seven fully slatted pens. Faecal samples were obtained at days 14, 36, 48, 60 and 70. A total of 770 238 raw sequences reads were generated by 454-pyrosequencing of the V3–V4 region of 16S rRNA gene (For additional information on experiments and methods, see Appendix 1). The sequences retained after different pre-processing steps ($n = 335\ 355$) were used for the reference-based operational taxonomic units (OTU) picking process (Table S1). Altogether, 99.8% of the sequences matched the GreenGenes reference core alignment set, while 0.2% were unknown (did not align to any sequence in the database above 75% percent of identity). The number of OTUs determined was 7857 (Table S2).

The dynamics of richness and diversity in the piglets and sows microbiota was computed with Chao1 estimator (Chao, 1984). At the level of 900 sequences per sample, the diversity was almost saturated for piglets at different ages, as revealed by the asymptotic shape of the sample rarefaction curves (Fig. S1). No significant difference in richness estimates was detected between piglets' gender (Fig. S1B) or cohousing animals, e.g. cohabitating piglets from multiple litters in the same pen (Fig. S1C).

The Firmicutes and Bacteroidetes phyla accounted for more than 90% of total sequences, similarly to previous findings in the ileal, caecal and faecal microbiota of weaning and finishing pigs (Poroyko *et al*., 2010; Kim *et al*., 2011; Schmidt *et al*., 2011; Buzoianu *et al*., 2012; Schokker *et al*., 2014). Other phyla were also present but at lower percentages (e.g. 5.14% *Proteobacteria*, 1.49% Spirochaetes, 0.76% Fusobacteria; Fig. S2). As shown in Fig. 1, piglet microbiota diversifies over the first weeks of life to create a homogeneous, rich and stable anaerobe-dominated microbial community after weaning (Katouli *et al*., 1997b; Jensen-Waern *et al*., 1998; Inoue *et al*., 2005; Thompson *et al*., 2008). These significant differences between suckling and weaned piglets were further confirmed by multivariate redundancy analysis of the Bray–Curtis distance (Fig. S3A), non-parametric multidimensional scaling (NMDS; Fig. S3B), network linked within a specified Jaccard distance (Fig. S3C and D; analysis of similarities (ANOSIM; Table S3)) and Random Forest, a supervised machine-learning technique (Table S4). Interestingly, *Bacteroides, Butyricimonas,* genera from the *Clostridiales* (e.g. *Oscillibacter, Clostridium sensu stricto, Clostridium IV, Clostridium XIVa)* and *Escherichia/Shigella* exhibited significant decline with increasing age [Fig. 1C; false discovery rate (FDR) < 0.05, Kruskal–Wallis test, Table S5]. The enrichment of *Bacteroides* and *Oscillibacter* genus in suckling pigs was in agreement with trends seen in humans (Palmer *et al*., 2007; Marcobal *et al*., 2011). It has been suggested that species from these genera are abundant in the neonate gastrointestinal microbiota because they are adapted to use wide range of both milk oligosaccharides and host-derived glycans (e.g. sulfomucin) as a unique carbon source (Palmer *et al*., 2007; Poroyko *et al*., 2010; Marcobal *et al*., 2011). Increased abundance of *Escherichia/Shigella* genus during lactation was also found in the faeces of humans (Maltby *et al*., 2013) and pigs (Konstantinov *et al*., 2006; Kim *et al*., 2011). Similar observations were recently observed in the pig gastrointestinal mucosa (Mann *et al*., 2014), indicating that pathobiont species are commonly present in the pig gastrointestinal tract, thus awaiting potential stressors to become pathogenic. At weaning, the sow's milk is totally replaced by cereal-based diets that have complex chemical composition (Spreeuwenberg *et al*., 2001), and piglets are separated from the sow and littermates. The introduction of a solid cereal-based diet, which in turn may modify the substrate availability and the physiological conditions of the gastrointestinal tract [e.g. fermentation products, luminal pH and bile acid concentration; (Opapeju *et al*., 2009; Kim *et al*., 2011; 2012)] was probably the main cause associated with the increased abundances of *Prevotella, Acetivibrio, Oribacterium, Paraprevotella, Roseburia* and *Succinivibrio* genera after weaning (Fig. 1C; FDR < 0.05; Kruskal–Wallis test; Table S5). A higher relative abundance of *Prevotella* (which represented $> 30\%$ of all 16S rRNA sequences after weaning) observed at weaning might have been due to the capacity of this genus to produce enzymes such as xylanases, mannanases, β-glucanases (Flint and Bayer,


A
B
R = 0.44 , P = 0.001
15000
10000
5000
0
Between
D14
D36
D48
D60
D70
Sow
Time
D14
D36
D48
D60
D70
Sow
sex
1
2
C
Relative abundance, %
60
40
20
0
Acetivibrio
Bacteroides
Barnesiella
Butyricimonas
Clostridium_IV
Clostridium_sensu_stricto
Clostridium_XIVa
Dialister
Escherichia/Shigella
Eubacterium
Lactobacillus
Mitsuokella
Oribacterium
Oscillibacter
Parabacteroides
Paraprevotella
Prevotella
Pseudoflavonifractor
Roseburia
Ruminococcus
Streptococcus
Succinivibrio
Treponema
unclassified_Porphyromonadaceae_
unclassified_Prevotellaceae_
unclassified_Lachnospiraceae
unclassified_Ruminococcaceae
D14
D36
D48
D60
D70
Sow

2008) that can degrade the polysaccharides in the cereal cell wall [e.g. arabinoxylan and cellulose; (Shah and Collins, 1990; Ivarsson *et al*., 2012)]. Yet, various confounding factors can hamper the interpretation and comparison of community shifts in pigs after weaning. Among these are gender effects, co-housing effect, genetic background and maternal effects. Here, we demonstrate that no significant association between microbiota composition and gender (FDR < 0.05; Mann–Whitney test; Table S6) or pen (FDR < 0.05; Kruskal–Wallis test; Table S7) was found throughout the experiment. Furthermore, in agreement with Buzoianu and colleagues (2012), who investigated transgenerational effects of feeding genetically modified maize to sows and their offspring on maternal and offspring intestinal microbiota from weaning to 115 days after birth, we did not find evidence that maternal microbiota is reflected in the offspring microbiota at weaning, or at later ages (Fig. 1A). Therefore, we suggest that stochastic factors might have a significant role in shaping the structure of the microbial communities within the first days of life (Schloss *et al*., 2012). However, a set of 315 OTUs (> 60% of the sequences) was shared between piglets and sows regardless of their age and gender, and identified as core (Table S8). Additionally, we found a total of 182 OTUs shared between mother and piglet pairs. Yet, out of 182 OTUs, only one was found in more than 90% of the mother–piglet pairs, being assigned to the sulfate-reducing bacteria *Desulfovibrio piger* (Fig. S4). *Desulfovibrio piger* is able to use the short chain fatty acids and sulfate released by other bacteria into the lumen during the initial stage of colonization (Fite *et al*., 2004; Rey *et al*., 2013). Other species identified that could be transferred from mothers to their piglet's intestines via faeces or breastfeeding were *Lactobacillus fermentum*, *Eubacterium coprostanoligenes* and *Clostridium ruminatum*. In humans, DNA of *L. fermentum* was previously identified in maternal faeces, breast milk and corresponding neonatal faeces within the same mother–neonate pair, which supports vertical transfer via breast milk (Jost *et al*., 2014). Interestingly, *L. fermentum* has been used as a growth-promoting feed supplement preventing and treating diarrhoea of weaned piglets and maximizing the average daily gain, crude protein apparent digestibility and serum specific IgG level (Yu *et al*., 2008). Thus, it is likely that in some cases, the lactobacilli from the mother's intestines or milk may colonize and adhere on the gastrointestinal tract epithelium of piglets, reinforcing their symbiotic relationship with the host and promoting their growth.

Evaluation of enterotype-like clusters in piglets

The temporal trajectory of bacterial communities in piglets showed that microbiota co-evolves with their hosts towards two different enterotype-like clusters, primarily distinguished by unclassified *Ruminococcaceae* and *Prevotella* levels (Fig. 2A). The optimal cluster number was found to be two by both the Calinski–Harabasz index as well as silhouette score (Fig. S5). Random Forest analyses confirmed distinct bacteria community signatures for piglets belonging to *Ruminococcaceae* or *Prevotella* clusters (baseline error = 0.194, cross-validation error = 0.011), and revealed the importance of *Prevotella* as the most discriminatory genus between the two clusters (Table S9). The relative abundances of 58 genera were significantly different between *Ruminococcaceae* and *Prevotella* clusters (FDR < 0.05; Mann–Whitney test; Table S10). The phylogenetic composition was highly similar to two of the enterotypes recently described in humans (Arumugam, Raes *et al*., 2011), and mice (Hildebrand *et al*., 2013). However, conversely to a fixed and discrete clustering of the gastrointestinal microbiota over time, we observed a dynamic enterotype-like clustering chiefly shifting by the relative abundance of the genus *Prevotella* across ages (Fig. S6). Whereas over the first 14 days of life, all piglets pertained to *Ruminococcaceae* cluster, weaning was associated with a shift in microbiota composition that moved 14 animals to *Prevotella* cluster. After weaning, we observed that 12 piglets were classified in the same cluster across ages, whereas 11 piglets remained in the same cluster in 80% of the time points. Lastly, eight piglets crossed the putative enterotype boundaries on a regular basis across time (Fig. S6), suggesting that discrete enterotype-like clusters are not fixed overtime for all animals. Similarly, Knights and colleagues (2014) projected a dense time series of 1 year's worth of daily gut microbiome samples and found that for some healthy subjects, enterotype can vary widely and continuously over time. In light of our findings, it

Fig. 1. Multidimensional reduction methods for elucidating diversity relationships of faeces microbiota in piglets and their mothers.
A. Genus-level taxonomic representation between piglets at different ages and their mothers linked within a specified Jaccard distance of 0.70. Two samples were considered 'connected' if the distance between them was less than 0.70. The relative position of points was optimized for the visual display of network properties. The point's shape indicates the gender; (B) Analysis of similarities (ANOSIM) function to test for differences in community composition among piglets at five age strata and their respective mothers. The analysis showed an R = 0.44 ($P < 0.001$), indicating that all samples within groups are more similar to each other than to any other samples from different groups.
C. Relative abundance of bacterial genera between piglets at different ages and their respective mothers. On the y-axis, relative abundance of 16S rRNA genes per sample and genera are shown. Genera names are coloured according to their phyla: Firmicutes (green blue), Bacteroidetes (orange), *Proteobacteria* (blue), Spirochaetes (dark blue). In all cases, day 14 (red, $n = 31$), day 36 (olive, $n = 31$), day 48 (green, $n = 31$), day 60 (green blue, $n = 31$), day 70 (dark blue, $n = 31$) and sows (pink, $n = 29$).

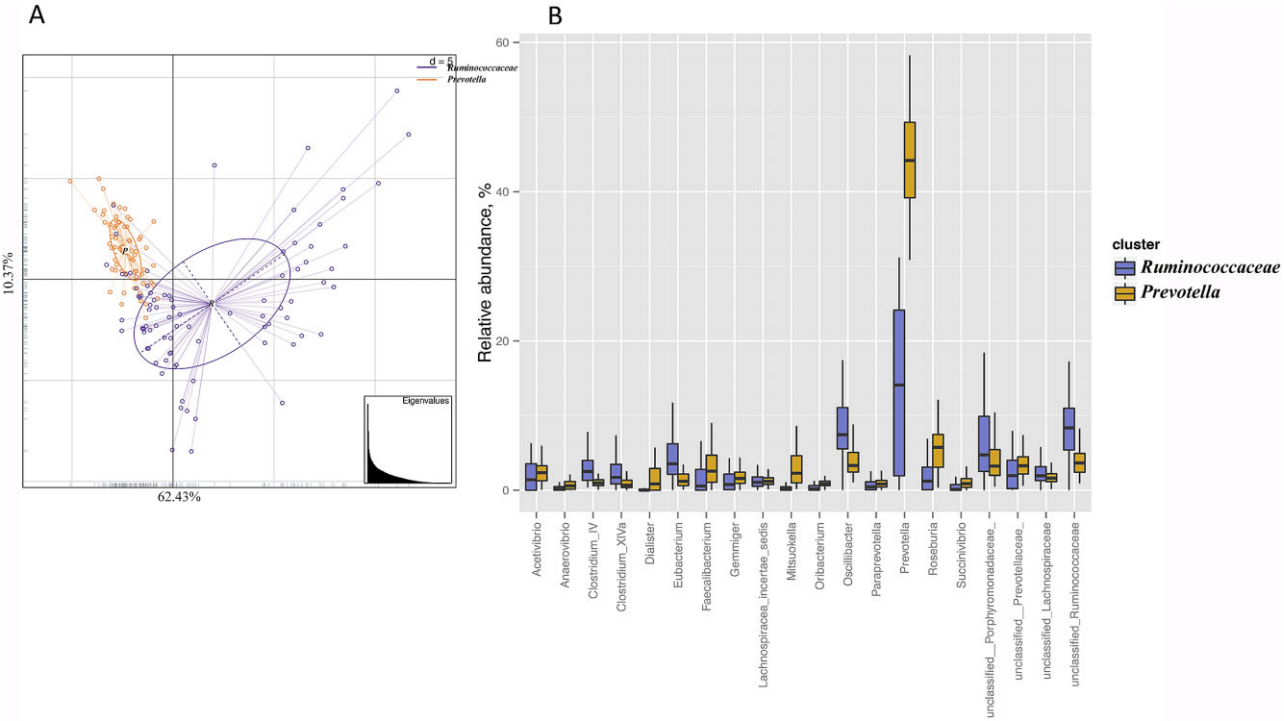


Fig. 2. Evaluation of enterotype-like clusters in piglets across five age strata.
A. Principal components analysis on a relative genus abundance matrix was performed from piglets' faeces. We used a probability distribution distance metric related to Jensen–Shannon divergence and PAM, as described by Arumugam, Raes and colleagues (2011). Individuals are represented by violet dots (*Ruminococcaceae* cluster) and orange dots (*Prevotella* cluster). Confidence ellipses around the centroids of the resulting clusters were used. The first axis extracted 62.43% of the variation, and the two axes kept 72.80% of the total inertia. Interclass PCA of genera profiles with enterotype-like clusters as instrumental variables was also assessed. Based on a Monte Carlo test with 999 replicates, a significant difference was found between the two clusters ($P < 0.001$); (B) Relative abundance of genera between the two enterotype-like clusters. Several genera were significantly different between clusters, with a cut-off of FDR < 0.05; Mann–Whitney test. On the y-axis relative abundance of 16S rRNA gene reads per sample and main genera are shown. Violet boxes: *Ruminococcaceae* cluster, and orange boxes: *Prevotella* cluster.

prompted the interest in discovering whether lactation-associated genera could be responsible for the shift of enterotype-like cluster after weaning and its stability over time. Although cause and effect is difficult to decipher, we identified three candidate biomarkers of community shifts after weaning. Lower abundance of clostridia such as *Oscillibacter* and *Clostridium* cluster XIVa but higher abundance of *Lactobacillus* genus in 14-day suckling piglets appeared to be discriminative ($P < 0.05$) and predisposed individuals to shift into *Prevotella* cluster after weaning. Interestingly, *Lactobacillus fermentum* abundance, which seems to be directly transmitted from mothers, was found to be two times higher in faeces of piglets belonging to the *Prevotella* cluster at any age, although their proportional representation was < 1% (age × cluster; $P < 0.001$).

Links between microbial communities, growth and secretory IgA levels: towards the identification of biomarkers

The average body weight (BW) of all 31 piglets at birth was homogeneous (1.65 ± 0.03 kg; 95% confidence interval of the mean was 1.51–1.78 kg; Fig. S7). However, the BW difference between the lightest and the heaviest piglets increased after birth (Fig. S7). In humans, numerous studies have revealed specific relationships between intestinal microbiota composition and abundance, and the host metabolism (Nicholson *et al.*, 2012; Le Chatelier *et al.*, 2013). Therefore, we first investigated whether the enterotype-like clustering might be related to performance and sIgA variation across ages. The enterotype-like clustering analysis showed that: (i) animals pertaining to *Ruminococcaceae* cluster presented better growth rates during lactation (Fig. 3) and (ii) animals belonging to *Prevotella* cluster presented lower growth rates during lactation but higher BW, and average daily gain (ADG) after weaning (Fig. 3). Similarly, analysis of luminal sIgA concentration at day 70 suggested a bimodal distribution, with piglets from *Prevotella* cluster having higher concentration of sIgA than the *Ruminococcaceae* cluster piglets (Fig. 4). We hypothesized that animals belonging to *Ruminococcaceae* cluster were more competitive during lactation because they presented higher abundance of *Bacteroides* and clostridia genera such as *Oscillibacter*

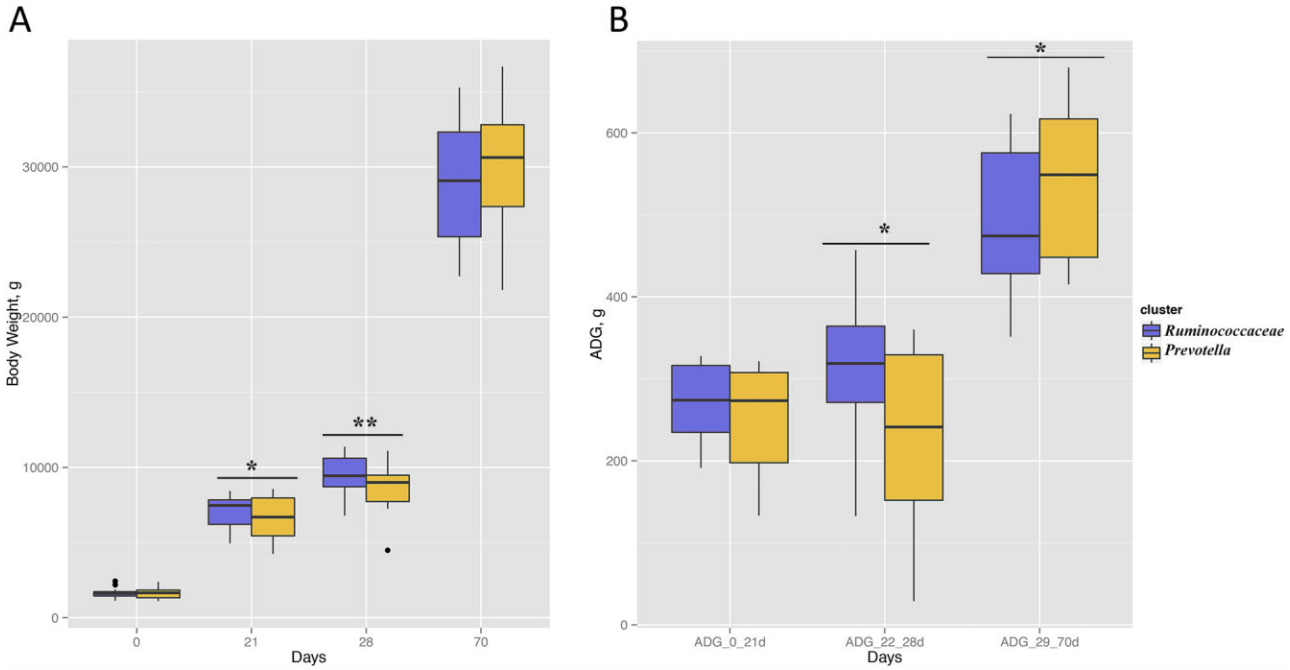


Fig. 3. Evaluation of growth performance distribution according to the two enterotype-like clusters in piglets at day 36.
A. The box plot graph represents the BW distribution across ages between the two enterotye-like clusters found at day 36; (B) The box plot graph represents the ADG distribution across ages between the two enterotye-like clusters at day 36. In all plots, individuals are represented by violet (*Ruminococcaceae* cluster) and orange (*Prevotella* cluster). *, **denote statistical significance at the 10%, 5% level respectively.

and *Clostridium XIVa*, which are able to digest the free milk oligosaccharides (Li *et al*., 2012), a major component of porcine milk (Tao *et al*., 2010) and favour the colonization by symbiotic anaerobes that have a positive effect on host performance. By contrast, animals belonging to *Prevotella* cluster presented better growth rates after weaning, suggesting that *Prevotella,* which plays an essential role in the process of complex dietary polysaccharides (Ellekilde *et al*., 2014), may promote increased uptake of monosaccharides in the host and confer performance advantage (McBurney and Sauer, 1993; Anguita *et al*., 2006). Additionally, the higher capacity of animals pertaining to *Prevotella* cluster to synthesize luminal sIgA may suggest that these individuals might

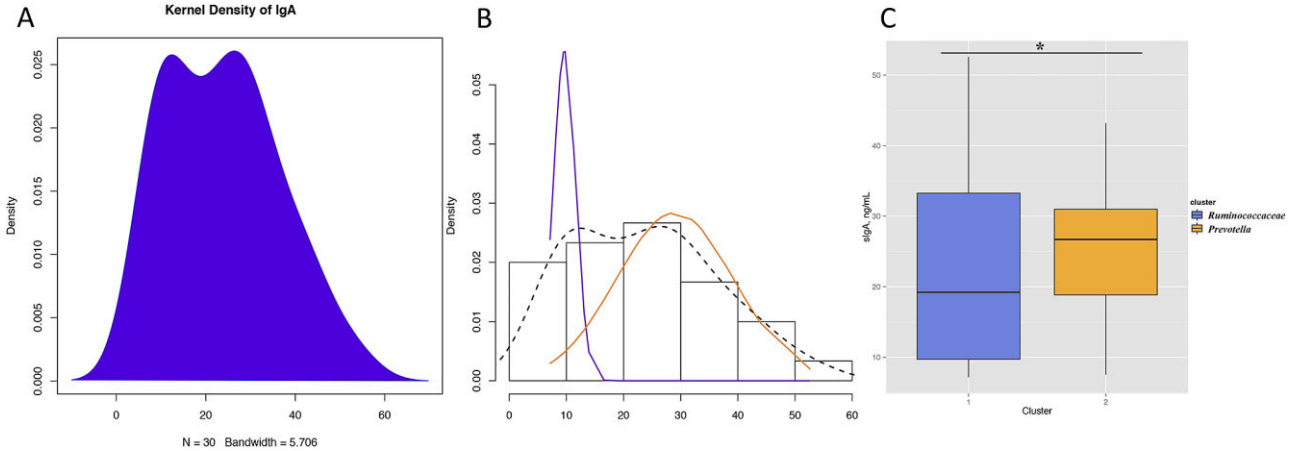


Fig. 4. sIgA concentration (ng/ml) in colon of 70 day piglets.
A. Kernel probability density function of sIgA secretion in the colon; (B) Probability density function of sIgA secretion in the colon using a mixture model. The black dotted curve represents the fitted mixture model, whereas coloured curves are scaled normal mixture components with means of 9.65 and 28.97 ng ml^{-1} respectively.
C. Distribution of sIgA concentration (ng/ml) in the colon according to the enterotype-like clusters in piglets at day 36. *denotes statistical significance at the 10% level.

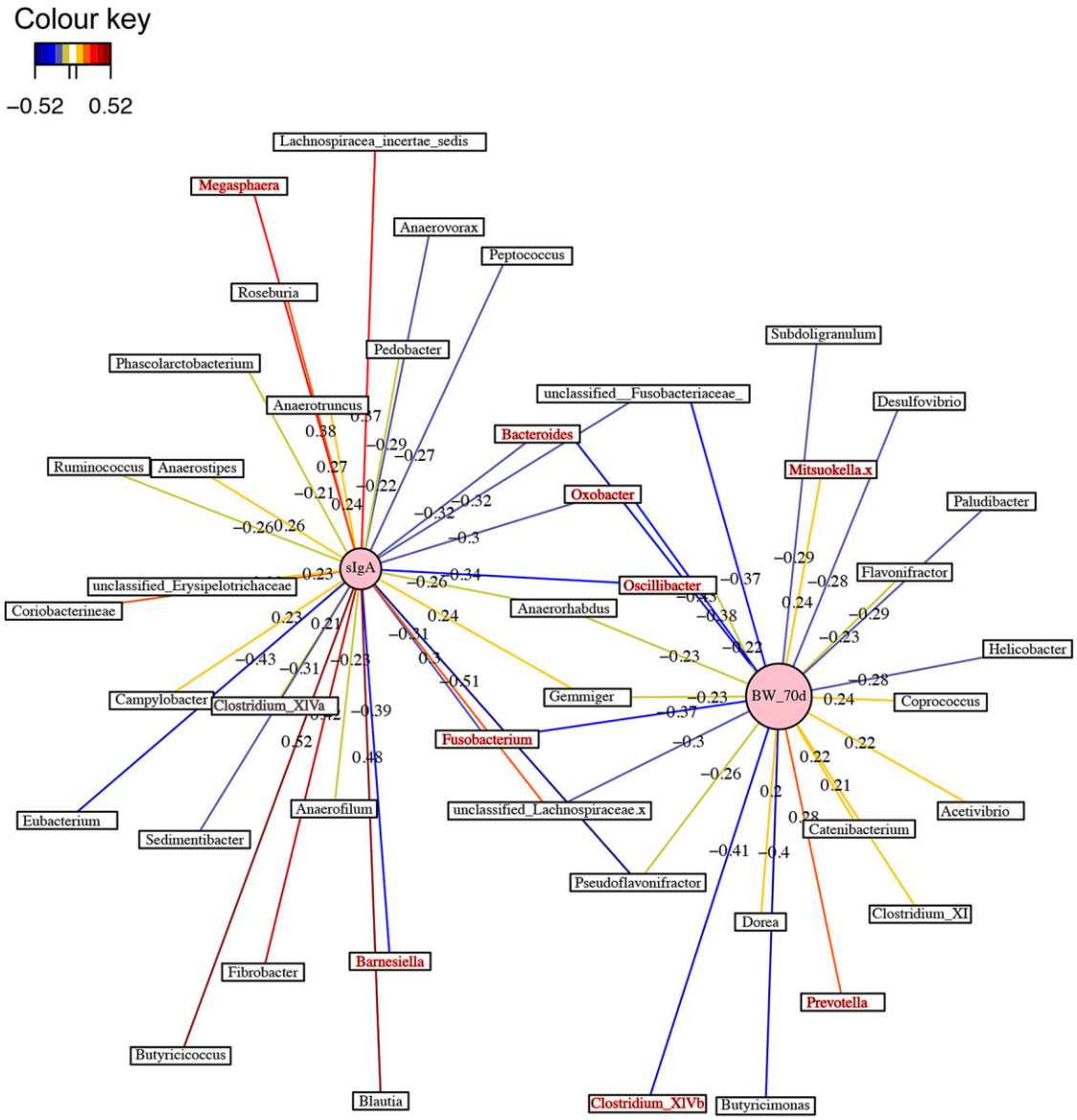


Fig. 5. Determination of potential key genera for body weight and sIgA production in piglets. Covariations between the relative abundances of bacterial general in the faecal microbiota at day 36 and different phenotypes at day 70 using sparse partial least squares regression. The most abundant genera were included in the analysis. The network is displayed graphically as nodes (genera and phenotypes) and edges (the biological relationship between nodes). The edge colour intensity indicates the level of the association: red = positive and blue = negative. Node shape indicates whether it is a phenotype (round) or genera (square). The score of the association was indicated under each edge. Pairwise associations with scores greater than 0.20 were projected.

have a decreased host's cost of responding to the intestinal microorganisms, and an additional performance efficiency. Because the individual's enterotype was variable after weaning for some animals and putative discrete clusters might be less effective to discover biomarkers than a predictive model constructed from the relative genera abundances (Knights *et al*., 2014), we tested whether the BW and ADG of piglets across ages, as well as sIgA concentration at day 70, could be influenced by the microbiota defined after weaning throughout Spearman correlation and sparse partial least squares analysis (sPLS) approaches (Le Cao *et al*., 2008a; 2011). This analysis identified a myriad of minor taxa, whose role in the pig intestine is not fully established. Some of them were highly associated with at least one phenotypic trait and could be considered as potential biomarkers (Fig. 5; Table S11). Increased abundance of *Fusobacterium* and clostridia (e.g. *Oxobacter*, *Oscillibacter)* at day 36 was negatively correlated with body weight and sIgA at day 70. The proportion of *Bacteroides* genus was also negatively correlated with the body weight, as seen elsewhere (Guo *et al*., 2008; Pedersen *et al*., 2013). Contrarily, higher

abundance of *Prevotella* and *Mitsuokella* were associated with increased BW. At the level of phyla, similar to Pedersen and colleagues (2013), a positive correlation between the weight gain and the relative abundance of Firmicutes was observed (Fig. S8). In mice, Hildebrandt and colleagues (2009) pointed to an association between alterations in energy intake and changes in gut microbiota such as increase in abundance of Firmicutes. Jumpertz and colleagues (2011) found that a 20% increase in abundance of Firmicutes resulted in higher energy harvest corresponding to approximately 150 kcal. Although the association between Bacteroidetes to Firmicutes ratio and performance phenotypes was not statistically significant (Fig. S8), we identified several bacterial genera belonging to both phyla Firmicutes and Bacteroidetes that may regulate weight gain and sIgA in pigs. As reported by Knights and colleagues (2014), the discovery of biomarkers in a supervised way, linking it directly to performance variables besides of relying uniquely on unsupervised clusters, provides promising information for understanding biological questions in the gastrointestinal tract for nutritional purposes and for genetic research. Taken together, our data suggest that gut microbiota might contribute to influence piglet growth performance in numerous ways, that could, consequently, be modulated.

Acknowledgements

This work was supported by the French National Research Agency (http://www.agence-nationale-recherche.fr/en/project-based-funding-to-advance-french-research/); project SusFlora: ANR-10-GENM-016). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We are grateful to F. Meurens, S. Vincent-Naulleau and N. Bertho for the animal experimental design and sampling management, and to S. Bouet, J. Lecardonnel, F. Andreoletti, P. Pinton, A. M. Cossalter, K. Le Roux for participating in the sample collection during the project.

References

Anguita, M., Canibe, N., Perez, J.F., and Jensen, B.B. (2006) Influence of the amount of dietary fiber on the available energy from hindgut fermentation in growing pigs: use of cannulated pigs and in vitro fermentation. *J Anim Sci* **84:** 2766–2778.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., *et al*. (2011) Enterotypes of the human gut microbiome. *Nature* **473:** 174–180.

Bearson, S.M., Allen, H.K., Bearson, B.L., Looft, T., Brunelle, B.W., Kich, J.D., *et al*. (2013) Profiling the gastrointestinal microbiota in response to Salmonella: low versus high Salmonella shedding in the natural porcine host. *Infect Genet Evol* **16:** 330–340.

Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D.S. (2009) Mixtools: an R package for analyzing finite mixture models. *J Stat Softw* **32:** 1–29.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001) Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125:** 279–284.

Benveniste, J., Lespinats, G., Adam, C., and Salomon, J.C. (1971) Immunoglobulins in intact, immunized, and contaminated axenic mice: study of serum IgA. *J Immunol* **107:** 1647–1655.

Buzoianu, S.G., Walsh, M.C., Rea, M.C., O'Sullivan, O., Cotter, P.D., Ross, R.P., *et al*. (2012) High-throughput sequence-based analysis of the intestinal microbiota of weanling pigs fed genetically modified MON810 maize expressing *Bacillus thuringiensis* Cry1Ab (Bt maize) for 31 days. *Appl Environ Microbiol* **78:** 4217–4224.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Chao, A. (1984) Nonparametric estimation of the number of classes in a population. *Scand J Stat* **11:** 256–270.

Clavel, T., Desmarchelier, C., Haller, D., Gerard, P., Rohn, S., Lepage, P., and Daniel, H. (2014) Intestinal microbiota in metabolic diseases: from bacterial community structure and functions to species of pathophysiological relevance. *Gut Microbes* **5:** 544–551.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., *et al*. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37:** D141–D145.

Cutler, D.R., Edwards, T.C., Jr, Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. (2007) Random forests for classification in ecology. *Ecology* **88:** 2783–2792.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al*. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

Ellekilde, M., Selfjord, E., Larsen, C.S., Jakesevic, M., Rune, I., Tranberg, B., *et al*. (2014) Transfer of gut microbiota from lean and obese mice to antibiotic-treated mice. *Sci Rep* **4:** 5922–5930.

Fite, A., Macfarlane, G.T., Cummings, J.H., Hopkins, M.J., Kong, S.C., Furrie, E., and Macfarlane, S. (2004) Identification and quantitation of mucosal and faecal desulfovibrios using real time polymerase chain reaction. *Gut* **53:** 523–529.

Flint, H.J., and Bayer, E.A. (2008) Plant cell wall breakdown by anaerobic microorganisms from the mammalian digestive tract. *Ann N Y Acad Sci* **1125:** 280–288.

Gibbons, R.D., Dorus, E., Ostrow, D.G., Pandey, G.N., Davis, J.M., and Levy, D.L. (1984) Mixture distributions in psychiatric research. *Biol Psychiatry* **19:** 935–961.

Guo, X., Xia, X., Tang, R., and Wang, K. (2008) Real-time PCR quantification of the predominant bacterial divisions in the distal gut of Meishan and Landrace pigs. *Anaerobe* **14:** 224–228.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., *et al*. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21:** 494–504.

Hamady, M., Lozupone, C., and Knight, R. (2010) Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* **4:** 17–27.

Hildebrand, F., Nguyen, T.L., Brinkman, B., Yunta, R.G., Cauwe, B., Vandenabeele, P., *et al*. (2013) Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol* **14:** R4.

Hildebrandt, M.A., Hoffmann, C., Sherrill-Mix, S.A., Keilbaugh, S.A., Hamady, M., Chen, Y.Y., *et al*. (2009) High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137:** 1716–1724, e1711–1712.

Inoue, R., Tsukahara, T., Nakanishi, N., and Ushida, K. (2005) Development of the intestinal microbiota in the piglet. *J Gen Appl Microbiol* **51:** 257–265.

Ivarsson, E., Liu, H.Y., Dicksved, J., Roos, S., and Lindberg, J.E. (2012) Impact of chicory inclusion in a cereal-based diet on digestibility, organ size and faecal microbiota in growing pigs. *Animal* **6:** 1077–1085.

Jensen-Waern, M., Melin, L., Lindberg, R., Johannisson, A., Petersson, L., and Wallgren, P. (1998) Dietary zinc oxide in weaned pigs – effects on performance, tissue concentrations, morphology, neutrophil functions and faecal microflora. *Res Vet Sci* **64:** 225–231.

Jost, T., Lacroix, C., Braegger, C.P., Rochat, F., and Chassard, C. (2014) Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environ Microbiol* **16:** 2891–2904.

Jumpertz, R., Le, D.S., Turnbaugh, P.J., Trinidad, C., Bogardus, C., Gordon, J.I., and Krakoff, J. (2011) Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am J Clin Nutr* **94:** 58–65.

Katouli, M., Foo, E., Kuhn, I., and Mollby, R. (1997a) Evaluation of the Phene Plate generalized microplate for metabolic fingerprinting and for measuring fermentative capacity of mixed bacterial populations. *J Appl Microbiol* **82:** 511–518.

Katouli, M., Lund, A., Wallgren, P., Kuhn, I., Soderlind, O., and Mollby, R. (1997b) Metabolic fingerprinting and fermentative capacity of the intestinal flora of pigs during pre- and post-weaning periods. *J Appl Microbiol* **83:** 147–154.

Kim, H.B., Borewicz, K., White, B.A., Singer, R.S., Sreevatsan, S., Tu, Z.J., and Isaacson, R.E. (2011) Longitudinal investigation of the age-related bacterial diversity in the feces of commercial pigs. *Vet Microbiol* **153:** 124–133.

Kim, H.B., Borewicz, K., White, B.A., Singer, R.S., Sreevatsan, S., Tu, Z.J., and Isaacson, R.E. (2012) Microbial shifts in the swine distal gut in response to the treatment with antimicrobial growth promoter, tylosin. *Proc Natl Acad Sci USA* **109:** 15485–15490.

Knights, D., Ward, T.L., McKinlay, C.E., Miller, H., Gonzalez, A., McDonald, D., and Knight, R. (2014) Rethinking 'enterotypes'. *Cell Host Microbe* **16:** 433–437.

Konstantinov, S.R., Awati, A.A., Williams, B.A., Miller, B.G., Jones, P., Stokes, C.R., *et al*. (2006) Post-natal development of the porcine microbiota composition and activities. *Environ Microbiol* **8:** 1191–1199.

Lalles, J.P., Bosi, P., Smidt, H., and Stokes, C.R. (2007) Nutritional management of gut health in pigs around weaning. *Proc Nutr Soc* **66:** 260–268.

Le Cao, K.A., Rossouw, D., Robert-Granie, C., and Besse, P. (2008a) A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol* **7:** 35–59.

Le Cao, K.A., Rossouw, D., Robert-Granie, C., and Besse, P. (2008b) A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol* **7:** 35–66.

Le Cao, K.A., Gonzalez, I., and Dejean, S. (2009a) integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics* **25:** 2855–2856.

Le Cao, K.A., Martin, P.G., Robert-Granie, C., and Besse, P. (2009b) Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics* **10:** 34.

Le Cao, K.A., Boitard, S., and Besse, P. (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12:** 253.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., *et al*. (2013) Richness of human gut microbiome correlates with metabolic markers. *Nature* **500:** 541–546.

Lepage, P., Seksik, P., Sutren, M., de la Cochetiere, M.F., Jian, R., Marteau, P., and Dore, J. (2005) Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm Bowel Dis* **11:** 473–480.

Levast, B., De Monte, M., Melo, S., Chevaleyre, C., Berri, M., Salmon, H., and Meurens, F. (2010) Differences in transcriptomic profile and IgA repertoire between jejunal and ileal Peyer's patches. *Dev Comp Immunol* **34:** 102–106.

Li, M., Bauer, L.L., Chen, X., Wang, M., Kuhlenschmidt, T.B., Kuhlenschmidt, M.S., *et al*. (2012) Microbial composition and in vitro fermentation patterns of human milk oligosaccharides and prebiotics differ between formula-fed and sow-reared piglets. *J Nutr* **142:** 681–689.

Li, W., and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22:** 1658–1659.

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. *Proc Natl Acad Sci USA* **109:** 1691–1696.

McBurney, M.I., and Sauer, W.C. (1993) Fiber and large bowel energy absorption: validation of the integrated ileostomy-fermentation model using pigs. *J Nutr* **123:** 721–727.

McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8:** e61217.

Macpherson, A.J. (2006) IgA adaptation to the presence of commensal bacteria in the intestine. *Curr Top Microbiol Immunol* **308:** 117–136.

Maltby, R., Leatham-Jensen, M.P., Gibson, T., Cohen, P.S., and Conway, T. (2013) Nutritional basis for colonization resistance by human commensal Escherichia coli strains HS and Nissle 1917 against E. coli O157:H7 in the mouse intestine. *PLoS ONE* **8:** e53957.

Mann, E., Schmitz-Esser, S., Zebeli, Q., Wagner, M., Ritzmann, M., and Metzler-Zebeli, B.U. (2014) Mucosa-associated bacterial microbiome of the gastrointestinal tract of weaned pigs and dynamics linked to dietary calcium-phosphorus. *PLoS ONE* **9:** e86950.

Marcobal, A., Barboza, M., Sonnenburg, E.D., Pudlo, N., Martens, E.C., Desai, P., *et al*. (2011) Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10:** 507–514.

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012) Host-gut microbiota metabolic interactions. *Science* **336:** 1262–1267.

Opapeju, F.O., Krause, D.O., Payne, R.L., Rademacher, M., and Nyachoti, C.M. (2009) Effect of dietary protein level on growth performance, indicators of enteric health, and gastrointestinal microbial ecology of weaned pigs induced with postweaning colibacillosis. *J Anim Sci* **87:** 2635–2643.

Palm, N.W., de Zoete, M.R., Cullen, T.W., Barry, N.A., Stefanowski, J., Hao, L., *et al*. (2014) Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158:** 1000–1010.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007) Development of the human infant intestinal microbiota. *PLoS Biol* **5:** e177.

Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013) Differential abundance analysis for microbial marker-gene surveys. *Nat Methods* **10:** 1200–1202.

Pedersen, R., Andersen, A.D., Molbak, 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 Microbiol* **13:** 30.

Poff, N.L., Olden, J.D., Merritt, D.M., and Pepin, D.M. (2007) Homogenization of regional river dynamics by dams and global biodiversity implications. *Proc Natl Acad Sci USA* **104:** 5732–5737.

Poroyko, V., White, J.R., Wang, M., Donovan, S., Alverdy, J., Liu, D.C., and Morowitz, M.J. (2010) Gut microbial gene expression in mother-fed and formula-fed piglets. *PLoS ONE* **5:** e12459.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5:** e9490.

Rey, F.E., Gonzalez, M.D., Cheng, J., Wu, M., Ahern, P.P., and Gordon, J.I. (2013) Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc Natl Acad Sci USA* **110:** 13582–13587.

Robinson, M.D., and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11:** R25.

Schloss, P.D., Schubert, A.M., Zackular, J.P., Iverson, K.D., Young, V.B., and Petrosino, J.F. (2012) Stabilization of the murine gut microbiome following weaning. *Gut Microbes* **3:** 383–393.

Schmidt, B., Mulder, I.E., Musk, C.C., Aminov, R.I., Lewis, M., Stokes, C.R., *et al*. (2011) Establishment of normal gut microbiota is compromised under excessive hygiene conditions. *PLoS ONE* **6:** e28284.

Schokker, D., Zhang, J., Zhang, L.L., Vastenhouw, S.A., Heilig, H.G., Smidt, H., *et al*. (2014) Early-life environmental variation affects intestinal microbiota and immune development in new-born piglets. *PLoS ONE* **9:** e100040.

Shah, H.N., and Collins, D.M. (1990) Prevotella, a new genus to include *Bacteroides melaninogenicus* and related species formerly classified in the genus Bacteroides. *Int J Syst Bacteriol* **40:** 205–208.

Shannon, C.E. (1997) The mathematical theory of communication. 1963. *MD Comput* **14:** 306–317.

Spor, A., Koren, O., and Ley, R. (2011) Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* **9:** 279–290.

Spreeuwenberg, M.A., Verdonk, J.M., Gaskins, H.R., and Verstegen, M.W. (2001) Small intestine epithelial barrier function is compromised in pigs with low feed intake at weaning. *J Nutr* **131:** 1520–1527.

Sutherland, D.B., and Fagarasan, S. (2012) IgA synthesis: a form of functional immune adaptation extending beyond gut. *Curr Opin Immunol* **24:** 261–268.

Tao, N., Ochonicky, K.L., German, J.B., Donovan, S.M., and Lebrilla, C.B. (2010) Structural determination and daily variations of porcine milk oligosaccharides. *J Agric Food Chem* **58:** 4653–4659.

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. *ISME J* **2:** 739–748.

Wilson, K.H., Blitchington, R.B., and Greene, R.C. (1990) Amplification of bacterial 16S ribosomal DNA with polymerase chain reaction. *J Clin Microbiol* **28:** 1942–1946.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., *et al*. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334:** 105–108.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., *et al*. (2012) Human gut microbiome viewed across age and geography. *Nature* **486:** 222–227.

Yu, H.F., Wang, A.N., Li, X.J., and Qiao, S.Y. (2008) Effect of viable Lactobacillus fermentum on the growth performance, nutrient digestibility and immunity of weaned pigs. *J Anim Feed Sci* **17:** 61–69.

Appendix 1. Experimental Procedures

Animals and sampling

Thirty-one large white piglets (15 females and 16 males) were studied from 29 commercial unrelated litters at French National Institute for Agronomical Research (INRA)'s experimental farm (Le Magneraud, France), together with their respective mothers ($n = 29$). The 31 piglets were selected carefully accounting for population–genetic structure, and covariates such as gender, and environmental influences (e.g. disease state, antimicrobials). One or two healthy piglets per mother were selected to avoid close genetic relationship between piglets within pens, which may cause difficulties to understand the individual variance underlying microbiota composition and would require the use of statistical models including

Table 1. Ingredient and chemical composition of the concentrates of post-weaning pigs from day 28 to day 70 and lactating sows.

	Post-weaning pigs	Lactating sows
Ingredients	% Dry matter feed	
Triticale	7.18	–
Wheat	20.51	5.84
Barley	15.38	20.13
Wheat starch	15.28	20.14
Corn	10.26	18.12
Forage peas	13.13	2.01
Soybean meal 48%	1.03	1.51
Palm oil	–	2.01
Beet pulp	2.36	5.03
Wheat middling's	–	5.03
Rapeseed meal	–	0.50
Sunflower meal	5.13	8.05
Canola oil	0.42	0.54
Dried brewers corn	0.63	0.53
Calcium carbonate	–	0.35
Dicalcium phosphate	1.03	1.01
Salt	0.22	0.06
Argyle	0.82	0.48
Liquid methionine	–	0.07
Lysine	0.17	0.51
Choline	1.23	–
Starter	7.18	–
Nutrients	% Dry matter feed	
Ash	5.8	5.92
Crude protein	17.49	16.00
Ether extract	3.80	4.62
NDF[a]	4.43	6.00
NFC[b]	68.48	67.46

a. NDF = neutral detergent fiber.
b. Non-fibrous carbohydrates = 100 minus the sum of ash, crude protein, NDF and fat.

random effects with a variance–covariance matrix dependent on the family structure of the piglets. Additionally, only one animal per litter was selected to prevent the maternal effects within the cohort of 31 piglets, which may hamper the interpretation and comparison of community shifts in pigs around weaning. All animal procedures were conducted according to the guidelines for the care and use of experimental animals established by INRA (ability for animal experimentation: A78-172, agreement for experimentation at le Magneraud: A-17661; protocol approved by a local ethics committee COMETHA Poitou-Charentes with the permit number: CE2013-2).

Piglets were weaned at 28 ± 0.2 days of age (±SEM) and weighted 9.1 ± 0.27 kg. After weaning, they were randomly assigned in seven different pens (fully slatted, temperature-controlled flat deck accommodation). Animals were kept in the same pen during the entire post-weaning period without the introduction of any new pigs. The management, environmental and housing factors have been controlled for all animals throughout the whole study. Creep feeding was provided during the last week of lactation. Post-weaning diets were formulated to exceed the INRA (1989) nutrient recommendations for monogastrics (Table 1). Cereals were used in the pig diet as the main sources of energy. Fresh, clean drinking water and diets were offered *ad libitum.* Piglets were individually weighed on day 0 (birth), day 21, day 28 (weaning) and day 70. Fresh faecal samples were obtained from the 31 piglets at days 14, 36, 48, 60 and 70. Piglets aged between 14 and 28 were individually placed on a sterile plastic tray, and faecal samples were collected while monitoring the piglets. At days 36, 48, 60 and 70, faecal samples were directly collected from the piglets' rectum. For sows, rectal samples were also onsite collected at day 14 following delivery. All faecal samples were directly frozen in liquid nitrogen and further stored at –80°C until use. On day 70, the animals were slaughtered and colonic luminal contents were sampled and stored at –80°C to assess the intestinal sIgA. None of the individuals (piglets or sows) received antibiotic therapy during the sampling period. Pigs were free of the principal swine infectious agents at the beginning and end of the study. Diarrhoea was not detected in pigs.

Luminal secretory IgA determination from luminal sample extracts

Colonic luminal contents were lyophilized overnight, and the dried faecal material was ground into fine powder. Two hundred milligrams of the resulting powder were dissolved in 2 ml of a cold extraction solution containing 100 mM phenylmethanesulfonyl fluoride, 5% bovine serum albumin and 0.1% sodium azide. The samples were thoroughly homogenized by a combination of manual shaking and mechanical homogenization on a vortex mixer during 30 s. The suspensions were then clarified by centrifugation at 2000 × *g* for 10 min at 4°C. The resulting supernatants were centrifuged once more at 10 000 × *g* for 10 min, and the supernatants obtained were transferred to 1.5 ml sterile Eppendorf tubes and stored at –20°C until use.

Luminal secretory IgA level was evaluated using pig immunoglobulin enzyme-linked immunosorbent assay (ELISA) quantification kit (Bethyl Laboratories, Montgomery, TX, USA) according to the manufacturers recommendations. Briefly, the Nunc Maxisorp plates (Thermo fisher Scientific, Roskilde, Denmark) were coated during 1 h at room temperature (RT) with 100 µlL of goat anti-porcine IgA antibody (A100-102A, Bethyl) diluted to 1:100 in 0.05 M carbonate-bicarbonate coating buffer, pH 9.6. Each well was washed four times with Tris-buffered saline solution (TBS) pH 8.0, containing 0.05% Tween 20 (TBS-Tw), and then blocked at RT for 30 min with TBS blocking buffer containing 1% BSA. The luminal extract supernatants and the provided standard pig serum provided were sequentially diluted with TBS-Tw buffer, 1% BSA and 100 µl were added to each well during

1 h at RT. The wells were washed and reacted with 100 μl/well of Horseradish Peroxidase-conjugated goat anti-pig IgA detection antibody (A100-102P, Bethyl) at RT during 1 h. After three washes, ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)) substrate (Thermo fisher Scientific, Rockford, IL, USA) was added to the plates, and the colour reaction was developed at room temperature for 15 min. The reaction was quantified spectrophotometrically at 414 nm by an Absorbance Microplate Reader (Labsystems Multiskan RC, Helsinki, Finland). Concentration of total IgA in intestinal lumen was calculated by including a pig reference serum with known IgA concentration, i.e. 650 μg ml^{-1} (RS10-107, Bethyl) as a standard that allowed to measure the amount of sIgA per gram of dried luminal content.

To evaluate the distribution of IgA concentration in the colon, a mixture model was fit to the IgA concentrations to provide an estimate of the number of populations or distributions that a set of data is drawn from (Gibbons *et al.*, 1984). The mix tools package (Benaglia *et al.*, 2009) in R was used, yielding a two-component mixture. The log-likelihood statistic of the bimodal mixture model was compared with the normal null model to test for a significant improvement in fit.

Composition, richness and diversity of faecal microbiota

Total DNA was extracted from aliquots of frozen faecal samples (200 mg; 155 samples at different age strata from 31 piglets and 29 mothers' samples), using a well-established method for the analysis of such ecosystem diversity (Lepage *et al.*, 2005), and microbiota composition was analysed by pyrosequencing of the V3–V4 [V3fwd: TACGGRAGGCAGCAG, V4rev: GGACTACCAGGGTATCTAAT; (Wilson *et al.*, 1990)] region of 16S rRNA gene. Polymerase chain reaction amplicons libraries were sequenced using the Roche 454 GS FLX Titanium platform. The resulting sequences ($n = 770\,238$) were analysed using the open source software package Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso *et al.*, 2010).

Following removal of the primers and barcodes, sequences were filtered as follows: (i) minimum and maximum read length of 300 bp and 500 bp respectively, (ii) no ambiguous base calls, (iii) no homopolymeric runs longer than 8 bp and (iv) minimum average Phred score > 27 within a sliding window of 50 bp. Chimeric sequences were curated using ChimeraSlayer (DeSantis *et al.*, 2006; Haas *et al.*, 2011). Sequences were aligned with NAST against the GreenGenes reference core alignment set (available in QIIME as core_set _aligned.fasta.imputed) using the 'align_seqs.py' script in QIIME. Sequences that did not cover this region at a percent identity $> 75\%$ were removed. Chimeric sequences were curated using ChimeraSlayer in QIIME (DeSantis *et al.*, 2006; Haas *et al.*, 2011). Operational taxonomic units were picked at a threshold of 97% similarity using cd-hit (Li and Godzik, 2006) from 'pick_otus.py' script in QIIME. Picking workflow in QIIME with the cd-hit clustering method currently involves collapsing identical reads using the longest sequence-first list removal algorithm, picking OTU and subsequently inflating the identical reads to recapture abundance information about the initial sequences. Singletons were removed, as only OTU that were present at the level of at least two reads in more than one sample were retained. The most abundant member of each OTU was selected through the 'pick_rep_set.py' script as the representative sequence. The resulting OTU representative sequences were assigned to different taxonomic levels (from phylum to genus) using the GreenGenes database (release August 2012), with consensus annotation from the Ribosomal Database Project naïve Bayesian classifier [RDP 10 database, version 6 (Cole *et al.*, 2009)]. To confirm the annotation, the resulting OTU representative sequences were then searched against the RDP database, using the online program SEQMATCH (http://rdp.cme.msu.edu/seqmatch/seqmatch_intro.jsp) and a threshold setting of 90% to assign a genus to each sequence.

Phylogenetic trees in the Newick format were produced in QIIME pipeline employing the FastTree program (Price *et al.*, 2010). To estimate a putative core OTU between different groups, we select the resulting OTU representative sequence at each group and their normalized abundance. To explore OTU that were ubiquitously present among groups, Venny, an interactive tool for the comparison of lists (http://bioinfogp.cnb.csic.es/tools/venny/index.html) was used.

The alpha diversity measurements were performed for $n = 100$ through 900 at intervals of 100 sequences (we chose 900 as the highest common subsample because the smallest sample had 944 sequences). The dynamics of richness and diversity in the piglets and sows microbiota was computed with Chao1 estimator (Chao, 1984) and the Shannon index (Shannon, 1997). The Chao1 estimates the number of taxa or functions present in a community, whereas the Shannon index of diversity takes into account both richness and evenness.

To estimate beta diversity measurements, which are a measure of separation of the phylogenetic structure of the OTU in one sample compared with all other samples, we first normalized the data to make taxonomic feature counts comparable across samples. Three different methods were separately evaluated to normalize the counts: (i) relative abundance normalization, (ii) the 'calcNormFactors' normalization function of edgeR package, which scales raw counts through a trimmed

mean of M-values (TMM) between samples (Robinson and Oshlack, 2010) and (iii) the cumulative-sum scaling (CSS) method (Paulson *et al.*, 2013), which divides raw counts by the cumulative sum of counts up to a percentile determined using a data-driven approach. The high reproducibility among normalization methods was validated by linear regression ('lm' function in the R statistical environment), with coefficient of correlations ≥ 0.78 between methods, and P values < 0.0001. In addition, we performed hierarchical clustering analysis to detect if any particular normalization method contributed largely to variability in the abundance, that is, whether it retained most of the information. The hierarchical clustering showed the remarkably high aggregation of three different tools, independently of the animal and time point, suggesting a similar robustness of different methods to normalize the abundance (Fig. S9). Since the genera abundance obtained using the three different methods was similar, we decided to perform beta diversity estimates measurements on the relative abundance normalized matrix. After normalization, OTU counts were binned into genus-level taxonomic groups, as well as family and phylum levels, according to the taxonomic assignments described previously. For Firmicutes/Bacteroidetes ratio, calculations were obtained for each individual using OTU counts.

Taxonomic changes across five age strata in piglets

The normalized OTU table combined with the phenotype metadata and phylogenetic tree comprised the data matrix used as input in phyloseq package in the R environment (http://www.r-project.org; version 3.0.1). From the otuSamTaxTree object created by phyloseq package, a sub otuSamTaxTree object that included only data of piglets was produced. Several distance metrics were considered in order to calculate the distance matrix of the different multidimensional reduction methods, including weighted/unweighted UniFrac distance (Hamady *et al.*, 2010) and non-phylogenetic distance metrics (Bray–Curtis, Jensen–Shannon divergence and Euclidian) using the phyloseq (McMurdie and Holmes, 2013) in R. Centred on genera taxa level, correspondence analysis was based on the Bray–Curtis distance measure, double principal coordinate analysis (DPCoA) and NMDS were based on different distance matrices. Network plots were calculated based on the Jaccard distance using the make_network function of phyloseq package at a 0.70 ecological distance. Two samples were considered 'connected' if the distance between them was less than 0.70. The relative position of points was optimized for the visual display of network properties. Lastly, redundancy analysis was implemented using the 'rda' function of the phyloseq package of R. The differences shown in redundancy analysis were assessed using Monte Carlo Permutation Procedure (999 replicates; 'randtest' function) of the Ade4 package in R. Additionally to multivariate analysis; we used the ANOSIM to test for intragroup dispersion. As specified by Poff and colleagues (2007), ANOSIM is a permutation-based test of the null hypothesis that within-group distances are not significantly smaller than between-group distances. The test statistic (R) can range from 1 to −1, with a value of 1 indicating that all samples within groups are more similar to each other than to any other samples from different groups. R is ≈ 0 when the null hypothesis is true, that distances within and between groups are the same on average. Furthermore, Kruskal–Wallis test was used to determine differential abundance of genera between ages. Results were corrected fro multiple testing using the Benjamini–Hochberg false discovery rate (q-value; Benjamini and colleagues (2001). *Post hoc* statistical testing for significant differences between all combinations of two groups was conducted using the Mann–Whitney test. To delineate the influence of gender and pen on microbiota composition across age, we used Mann–Whitney test and Kruskal–Wallis test, respectively, followed by multiple testing correction. Moreover, we applied random forest machine-learning analyses to identify bacterial genera level that differentiates faecal community composition between groups. The purpose of a classifier such as random forest is to learn a function that maps a set of input values or predictors (here, relative to genera abundances) to a discrete output value (here, relative to all age group combinations) (Yatsunenko *et al.*, 2012). Random forest is a powerful classifier that can use nonlinear relationships and complex dependencies between taxa. The degree of the success of the method is its ability to classify unseen samples correctly, estimated by training it on a subset of samples, and using it to categorize the remaining samples (Cutler *et al.*, 2007). The cross-validation error is compared with the baseline error that would be achieved by always predicting the most common category. Additionally, random forest assigns an importance score to each genus by estimating the increase in error caused by eliminating that genus from the set of predictors. It also reports the Gini importance of a variable, which is computed as the sum of the Gini impurity decrease of every node in the forest for which that variable was used for splitting (Cutler *et al.*, 2007). All error estimates and genera importance scores were averaged over 100 rarefactions at the same sample size for each community to control for sequencing effort.

Detection of genera shared by mother–piglet pairs that might represent vertical inheritance

We obtained faecal samples from 29 mother–piglet pairs at 14 days after delivery. Sows had an average of 12.35 ± 2.09 pigs born alive per litter, 0.89 ± 0.95 piglets

that died after birth and 0 mummified foetuses per litter. Piglets averaged 1.89 ± 0.43 kg body weight at birth. One mother–infant pair had to be discarded because of sow health problems.

To characterize whether bacteria transmitted directly from mothers were able to occupy the piglet gastrointestinal niches and persist as lactation-adapted microbiota, we characterized the shared OTUs between piglets and their corresponding maternal faeces. Additionally, the core microbiome between piglets and mothers was established using Venny (http://bioinfogp.cnb.csic.es/tools/venny/index.html). We considered an OTU being a member of the core microbiome if it was present within all subjects (100%) sampled.

Evaluation of enterotype-like clusters in piglets

Principal component analysis (PCA) and clustering analysis were computed on different distance matrix to document the presence of enterotype-like clusters in piglets. The data set that included only data of piglets was considered. Clustering analysis was performed following the original criteria reported by Arumugam and colleagues (2011), who used the maximization of the Calinski–Harabasz index and the silhouette index (Wu *et al*., 2011) to select the optimal number of clusters. The following clustering algorithms implemented in R packages were used: via partitioning around k-medoids algorithm (PAM clustering), k-means and hierarchical clustering using different linkages. Interclass PCA of faeces genera composition with enterotype-like clusters as instrumental variable was also assessed, based on a Monte Carlo test with 999 replicates. Lastly, random forest analysis was also performed to identify bacterial genera that differentiated faecal community composition between the two identified enterotype-like clusters. Random forest analysis was performed for each comparison on 500 rarefied versions of the data.

Determination of the potential genera biomarkers for growth and sIgA production in piglets

Two different statistical approaches were applied to find possible associations between the faeces microbiota composition in piglets and the sIgA concentrations and performance, and to detect which of these genera could be considered as potential genera biomarkers: (i) mixed-effects ANOVA model to delineate whether there was a significant difference between the average values of phenotype traits for the two different enterotype-like clusters and (ii) the sPLS and the non-parametric Spearman rank correlation to link genera relative abundances directly to phenotype traits. The statistical mixed-effects ANOVA model included the enterotype-cluster type as fixed effect, and piglet nested within pen as a random effect to account for any potential dependencies between animals within pen. A significance level of $P < 0.05$ was accepted. The sPLS maximized the covariance between two data sets by searching for linear combinations of the variables. Furthermore, it imposed sparsity within the context of partial least squares and thereby carried out dimension reduction and variable selection simultaneously (Le Cao *et al*., 2008a; 2011). To evaluate the statistical significance of covariation between the genera proportion and the distinct phenotypes, we performed the M-fold or leave-one-out cross-validation, estimating the mean squared error of prediction (MSEP), the R^2 and Q^2 for each phenotype in the data set. An X variable contributed significantly to the prediction if $Q^2_h \geq (1–0.95^2) = 0.0975$ (Le Cao *et al*., 2008b). The mixOmics package in R was used to carry out sPLS analyses (Le Cao *et al*., 2009a,b). We used the 'network' functions to generate the images from sPLS. The 'network' function calculated a similarity measure between X and Y variables in a pairwise manner. The output was a graph in which each X-and Y-variable corresponds to a node, and the edges included in the graph display the associations between the nodes. Lastly, non-parametric Spearman rank correlation and Pearson correlation were calculated between performance variables and sIgA using the corrplot package in R.

Data submission

The 16S rRNA sequences described in the study were deposited at DDBJ/EMBL/GenBank under accession n° KP101623 – KP109479. The version described in this paper belongs to National Center for Biotechnology Information (NCBI) BioProject PRJNA266269 and NCBI BioSample SAMN03160604.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Microbiota richness and diversity in piglets and sows.
A. Estimation of the abundance of unique OTU in piglets at different ages and in sows using Chao1 index (Chao, 1984). Data were based on 900 sequences per sample. The values are means, and colour bars indicate the 95% confidence intervals.
B. Estimation of the abundance of OTUs between genders in piglets; (C) Estimation of the abundance of OTUs between the seven fully slatted flat deck accommodation where piglets were accommodated.
Fig. S2. Relative abundance of phyla in faeces of pigs among five age strata and sows.
The 16S rRNA gene V3–V4 sequences were binned into OTUs, normalized and summarized by phylum. Phyla were

colour coded according to the scheme on the right. The horizontal axis displays sample collection time points.

Fig. S3. Multidimensional reduction methods for elucidating diversity relationships of faeces microbiota in piglets across ages.

A. Redundancy analysis (RDA) analysis of Bray–Curtis distances to compare faecal communities at the level of genera that differ between ages. Both PC axes 1 and 2 were plotted. Together they explained 58% of whole variation.

B. Non-parametric multidimensional scaling (NMDS) ordination graphic generated by weighted unifrac MDS. The NMDS coordinates are generated with the weighted unifrac distance matrix as the argument, and two dimensions specified by default. The fit statistics for observations was 0.15.

C and D. Network diagram of genera between different age strata (nodes) that were linked within a specified Jaccard distance. The network was calculated using the make_network function of phyloseq package at a 0.70 ecological distance. Two samples were considered 'connected' if the distance between them was less than 0.70. The relative position of points was optimized for the visual display of network properties. The point's shape indicates the gender (C) or pen (D). In all plots, day 14 (red, $n = 31$), day 36 (olive, $n = 31$), day 48 (green, $n = 31$), day 60 (blue, $n = 31$) and day 70 (violet, $n = 31$).

Fig. S4. Operational taxonomic units showing the largest presence within the mother–piglet pairs.

The cell plot appears with columns indicating the mother–piglet pair, and rows the absence/presence of the OTUs found in more than 75% of the mother–piglet pairs. One mother–infant pair had to be discarded because of sow health problems. Green colour is assigned to presence, whereas white to absence.

Fig. S5. Evaluation of enterotype-like clusters in piglets across five age strata.

A. The number and quality of clusters were validated by maximizing the silhouette index as described by Wu and colleagues (2011). Several clustering algorithms, including complete hierarchical clustering, k-means clustering and partitioning around medoids (PAM) were tested.

B. The number and quality of clusters were validated using the Calinski–Harabasz (CH) Index as described by Arumugam, Raes and colleagues (2011), which showed good performance in recovering two clusters.

Fig. S6. Dynamics of enterotype-like clusters in piglets across five age strata.

In the plot, violet cells: *Ruminococcaceae* cluster, and orange cells: *Prevotella* cluster.

Fig. S7. Body weight distribution across ages.

The bar plot graph represents the body weight distribution at birth, 21 days, 28 days and 70 days of all 31 piglets respectively.

Fig. S8. Correlation between Firmicutes/Bacteroidetes ratio at day 36, performance variables and sIgA in piglets. The heat map visualizes all Spearman correlation coefficients between the Firmicutes/Bacteroidetes ratio (F : B ratio), performance variables (BW_70d: body weight at 70 days; ADG_29_70d: average daily gain between days 29 and 70) and sIgA (IgA). Each row and column represents a single trait. The negative correlations between traits are indicated in blue and the positive correlations in red.

Fig. S9. Hierarchical clustering to assess the capabilities of three different normalization methods.

Three different normalization methods have been tested: (i) relative abundance normalization; which divides raw counts from a particular sample by the total number of reads in each sample; (ii) the 'calcNormFactors' normalization function of edgeR package, which scales raw counts through a TMM between samples (Robinson and Oshlack, 2010); and (iii) the CSS) method (Paulson *et al.*, 2013), which divides raw counts by the cumulative sum of counts up to a percentile determined using a data-driven approach. A hierarchical clustering analysis matrix is represented (distance = correlation; aggregation method = ward). In the tree, each unit corresponds to one animal, time point and normalization method.

Table S1. Summary of study samples and faecal bacterial 16S rRNA gene amplicon sequence datasets.

Sequences reads were generated by 454-pyrosequencing of the V3–V4 region of 16S rRNA gene and analysed using the open source software package Quantitative Insights Into Microbial Ecology (QIIME) (Caporaso *et al.*, 2010). After trimming the primers and barcodes, the sequences were filtered and clustered at a threshold of 97% similarity level using cd-hit (Cluster Database at High Identity with Tolerance; Li and Godzik, 2006). The most abundant member of each OTU was selected as the representative sequence and assigned to different taxonomic levels using the GreenGenes taxonomy database, with consensus annotation from the RDP naïve Bayesian classifier (Cole *et al.*, 2009). Subject description: Day 14 (D14, $n = 31$), Day 36 (D36, $n = 31$), Day 48 (D48, $n = 31$), Day 60 (D60, $n = 31$), day 70 (D70, $n = 31$) and sow ($n = 29$).

Table S2. The OTU abundance in piglets at different ages and in sows.

The OTU abundance is listed by group including sequence number, abundance and closest reference strain and similarity. The OTU sequences were assigned to different taxonomic levels (from phylum to genus) using the GreenGenes database (release August 2012), with consensus annotation from the Ribosomal Database Project naïve Bayesian classifier [RDP 10 database, version 6 (Cole *et al.*, 2009)]. The dominant OTU in piglets at day 14 (7.3% of reads, corresponding to *Bacteroides fragilis*) was only detected with less than two sequences in piglets at all subsequent time points and in sows.

Table S3. Results from analysis of similarities (ANOSIM) test between faecal microbiota of piglets across age strata and sows.

ANOSIM is a permutation-based test where the null hypothesis states that within-group distances are not significantly smaller than between-group distances. The test statistic (R) can range from 1 to −1, with a value of 1 indicating that all samples within groups are more similar to each other than to any other samples from different groups. Inter-individual variation in microbiota communities during the first 14 days of life were significantly higher than that observed at older ages and for the sows. Mean ± 95% CI and R values are shown.

Table S4. Random forest machine-learning analyses discriminates the faecal microbiota according to age strata.

Random Forest confirmed distinct community signatures before and after weaning (baseline error = 0.348, cross-validation error = 0.00001). The higher values of mean

decrease in accuracy and the mean decrease in Gini indexes were found for *Prevotella*, confirming the importance of this genus as a discriminatory taxon for piglets before and after weaning. Main contributors are indicated in bold characters.

Table S5. Differences in relative abundance of genera between age strata.

We performed a Kruskal–Wallis test to determine differential abundances of genera between ages with FDR multiple correction. FDR < 0.05 were considered significant. The subject numbers (#N) are counts of subjects that contain corresponding genera in each age strata faecal samples. The average percentage of each genus is indicated in the column '% Total'.

Table S6. Influence of gender on microbiota composition across age.

We performed a Mann–Whitney test to determine differential abundance of genera between genders (male versus female) across ages with FDR multiple correction. The subject numbers (#N) are counts of subjects that contain corresponding genera in each age strata faecal samples. The average percentage of each genus is indicated in the column '% Total'.

Table S7. Influence of pen on microbiota composition across ages.

We performed a Kruskal–Wallis test to determine differential abundances of genera between seven different pens across ages with FDR multiple correction. The subject numbers (#N) are counts of subjects that contain corresponding genera in each age strata faecal samples. The average percentage of each genus is indicated in the column '% Total'.

Table S8. The core OTU abundance in piglets at different age and sows.

An OTU was considered a member of a core if presented in all subjects (100%) sampled. The OTU abundance is listed by group including sequence number, abundance and closest reference strain and similarity.

Table S9. Random Forest classifier discriminates the faecal microbiota according to enterotype-like clustering.

Random Forest confirmed distinct community signatures before and after weaning (baseline error = 0.194, cross-validation error = 0.011). The higher values of mean decrease in accuracy and the mean decrease in Gini indexes were found for *Prevotella*, revealing the importance of *Prevotella* as the most discriminatory genus between the two enterotype-like clusters. Main contributors are indicated in bold characters.

Table S10. The Mann–Whitney test results between the two enterotype-like clusters.

We performed a Mann–Whitney test to determine differential abundances of genera between the two enterotype-like clusters with FDR multiple correction. FDR < 0.05 were considered significant.

Table S11. Determination of potential genera biomarkers for performance and immune system response in piglets.

Correlation between relative abundance of microbiota at day 36 and different phenotypes using non-parametric Spearman rank correlation.