



**HAL**  
open science

## A reference gene catalogue of the pig gut microbiome

Liang Xiao, Jordi Estellé, Pia Kiilerich, Yulixis Ramayo Caldas, Zhongkui Xia, Qiang Feng, Suisha Liang, Niels Jørgen Kjeldsen, Chuan Liu, Emmanuelle Maguin, et al.

► **To cite this version:**

Liang Xiao, Jordi Estellé, Pia Kiilerich, Yulixis Ramayo Caldas, Zhongkui Xia, et al.. A reference gene catalogue of the pig gut microbiome. *Nature Microbiology*, 2016, 1, pp.1-6. 10.1038/nmicrobiol.2016.161 . hal-01607746

**HAL Id: hal-01607746**

**<https://hal.science/hal-01607746v1>**

Submitted on 28 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# A reference gene catalogue of the pig gut microbiome

Liang Xiao<sup>1†</sup>, Jordi Estellé<sup>2†</sup>, Pia Kiilerich<sup>3†</sup>, Yuliaxis Ramayo-Caldas<sup>2</sup>, Zhongkui Xia<sup>1</sup>, Qiang Feng<sup>1‡</sup>, Suisha Liang<sup>1</sup>, Anni Øyan Pedersen<sup>4</sup>, Niels Jørgen Kjeldsen<sup>4</sup>, Chuan Liu<sup>1,5</sup>, Emmanuelle Maguin<sup>6</sup>, Joël Doré<sup>6,7</sup>, Nicolas Pons<sup>7</sup>, Emmanuelle Le Chatelier<sup>7</sup>, Edi Prifti<sup>7‡</sup>, Junhua Li<sup>1,8</sup>, Huijue Jia<sup>1</sup>, Xin Liu<sup>1</sup>, Xun Xu<sup>1</sup>, Stanislav D. Ehrlich<sup>7,9</sup>, Lise Madsen<sup>1,3,10</sup>, Karsten Kristiansen<sup>1,3\*</sup>, Claire Rogel-Gaillard<sup>2\*</sup> and Jun Wang<sup>1,3\*</sup>

**The pig is a major species for livestock production and is also extensively used as the preferred model species for analyses of a wide range of human physiological functions and diseases<sup>1</sup>. The importance of the gut microbiota in complementing the physiology and genome of the host is now well recognized<sup>2</sup>. Knowledge of the functional interplay between the gut microbiota and host physiology in humans has been advanced by the human gut reference catalogue<sup>3,4</sup>. Thus, establishment of a comprehensive pig gut microbiome gene reference catalogue constitutes a logical continuation of the recently published pig genome<sup>5</sup>. By deep metagenome sequencing of faecal DNA from 287 pigs, we identified 7.7 million non-redundant genes representing 719 metagenomic species. Of the functional pathways found in the human catalogue, 96% are present in the pig catalogue, supporting the potential use of pigs for biomedical research. We show that sex, age and host genetics are likely to influence the pig gut microbiome. Analysis of the prevalence of antibiotic resistance genes demonstrated the effect of eliminating antibiotics from animal diets and thereby reducing the risk of spreading antibiotic resistance associated with farming systems.**

Domesticated pigs are a major food source worldwide and thus important in global food security, but large-scale pig farming faces environmental challenges. Increasing the understanding of the interaction between the gut microbiota and the host has the potential to facilitate knowledge-based development of sustainable pig production by increasing feed efficiency and general health in pig farming. Furthermore, the spread of antibiotic-resistant bacteria and genes is a matter of great global concern, with legislation and research in this area being important issues<sup>6–11</sup>. Hence, an evaluation of the impact of prohibiting the use of antibiotics as growth promotants is needed.

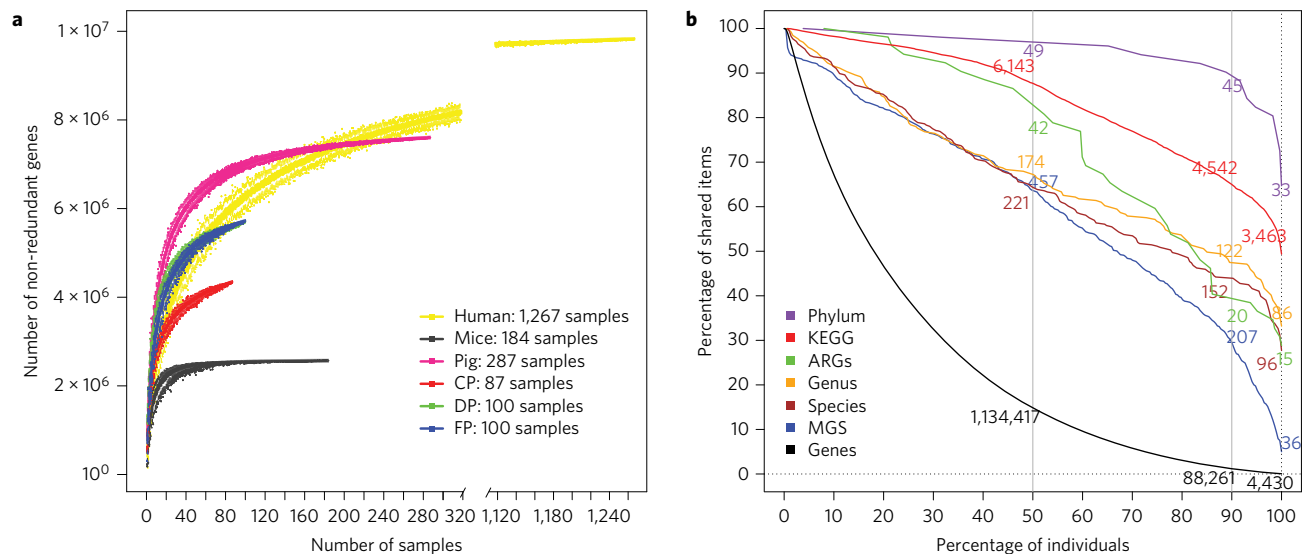
## Results

We collected 287 faecal samples from France (100 pigs), Denmark (100 pigs) and China (87 pigs), from 17 breeds or selected lines from 11 farms. The pigs varied in age, sex and antibiotic supplementation (Supplementary Table 1). Deep sequencing of faecal DNA samples

generated 1,758 Gb of high-quality data with an average of 6.13 Gb per sample (Supplementary Table 2), and allowed us to identify 7,685,872 non-redundant (NR) genes with an average N50 contig length of 1.89 kb (Supplementary Tables 2 and 3). A rarefaction analysis including all samples revealed a curve approaching saturation (Fig. 1a), and the chao2 index indicated that we had captured 99.9% of microbial genes in the samples. Half of the NR genes could be taxonomically classified. Of these, more than 98% could be assigned to the Bacteria super kingdom (Supplementary Fig. 1), the remaining 2% being genes from archaea and eukaryotes. At the phylum level, most of the annotated genes (28.73%) belonged to Firmicutes, followed by Bacteroidetes (9.28%). Only 7.6% and 0.33% of the genes could be annotated to specific bacterial genera and species, respectively, reflecting the paucity of sequenced pig gut bacterial genomes. By contrast, 16.3% of genes of the integrated human reference catalogue could be assigned to a bacterial genus<sup>4</sup>. At the genus level, most of the annotated genes (1.90%) belonged to *Prevotella*, followed by *Bacteroides* (0.80%), *Clostridium* (0.79%), *Ruminococcus* (0.72%) and *Eubacterium* (0.51%). Of the 719 annotated metagenomic species (MGS), 497 MGS could be annotated to known bacteria, with the remaining 222 MGS representing hitherto unknown microbial entities. A total of 353 MGS could be assigned to a phylum, but only 33 and one could be assigned to a genus and a species, respectively (Supplementary Fig. 1b). Among the MGS, *Prevotella* was the bacterial genera to which most MGS (0.97%) could be annotated, followed by *Ruminococcus* (0.55%), *Eubacterium* (0.42%), *Lactobacillus* (0.42%) and *Helicobacter* (0.41%). Functional classification of annotated NR genes and MGS using KEGG (Supplementary Fig. 2a) and eggNOG (Supplementary Fig. 2b) revealed a predominance of pathways related to genetic information processing (replication and repair), metabolism (carbohydrates, amino acids and nucleotides) and to information processing (membrane transport). We identified a common set of 4,430 NR genes, 36 MGS and 3,463 related annotated functions shared by 100% of the 287 pig samples, suggesting the existence of a core of genes, species and functions in the gut microbiome in pigs (Fig. 1b and Supplementary Fig. 3). With a

<sup>1</sup>BGI-Shenzhen, 518083 Shenzhen, China. <sup>2</sup>GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. <sup>3</sup>Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark. <sup>4</sup>SEGES Pig Research Centre, DK-1609 Copenhagen V, Denmark. <sup>5</sup>Shenzhen Engineering Laboratory of Detection and Intervention of Human Intestinal Microbiome. <sup>6</sup>MICALIS Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France. <sup>7</sup>MetaGénoPolis, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France. <sup>8</sup>Shenzhen Key Laboratory of Human Commensal Microorganisms and Health Research. <sup>9</sup>King's College London, Centre for Host-Microbiome Interactions, Dental Institute Central Office, Guy's Hospital, London SE1 9RT, UK. <sup>10</sup>National Institute of Nutrition and Seafood Research (NIFES), Postboks 2029, Nordnes, N-5817 Bergen, Norway. <sup>†</sup>Present addresses: Department of Human Microbiome, School of Stomatology, Shandong University, Shandong Provincial Key Laboratory of Oral Tissue Regeneration, Jinan, China. (Q.F.). Institute of Cardiometabolism and Nutrition, 75013 Paris, France. (E.P.). <sup>‡</sup>These authors contributed equally to this work.

\*e-mail: [claire.rogel-gaillard@jouy.inra.fr](mailto:claire.rogel-gaillard@jouy.inra.fr); [kk@bio.ku.dk](mailto:kk@bio.ku.dk); [wangjun30@gmail.com](mailto:wangjun30@gmail.com)



**Figure 1 | Rarefaction curves and shared metagenome among pigs.** **a**, Rarefaction curves of the pig (pink), human (yellow) and mouse (grey) samples generated from the NR gene counts. The rarefaction curves drawn from the Chinese (CP red), French (FP blue) and Danish (DP green) pigs and from the whole set of 287 pigs confirm the complementarities of the three animal sets for providing good coverage of the gene catalogue of the pig gut microbiome. **b**, Number of shared microbiome features among pigs at different frequency thresholds for the number of genes (black), phyla (purple), genera (orange), species (brown), metagenomic species (MGS, blue), antibiotic resistance genes (ARGs, green) and KEGG pathways (red). The percentages of shared items and animals are represented on the y and x axes, respectively. The absolute numbers for each item are indicated at the intercept between the percentages of items and animals at the thresholds of 50, 90 and 100%. Fewer than 20% of genes (1,134,417 genes) were shared by 50% of the pigs, while about 90% of the KO functions (6,143) were shared by 50% of the pigs, suggesting redundancy of genes for similar functions.

cutoff requiring that a core should be shared by 90% of the pigs, the common sets increased to 88,261 NR genes, 207 MGS and 4,542 KEGG orthology (KO) functions (Fig. 1b).

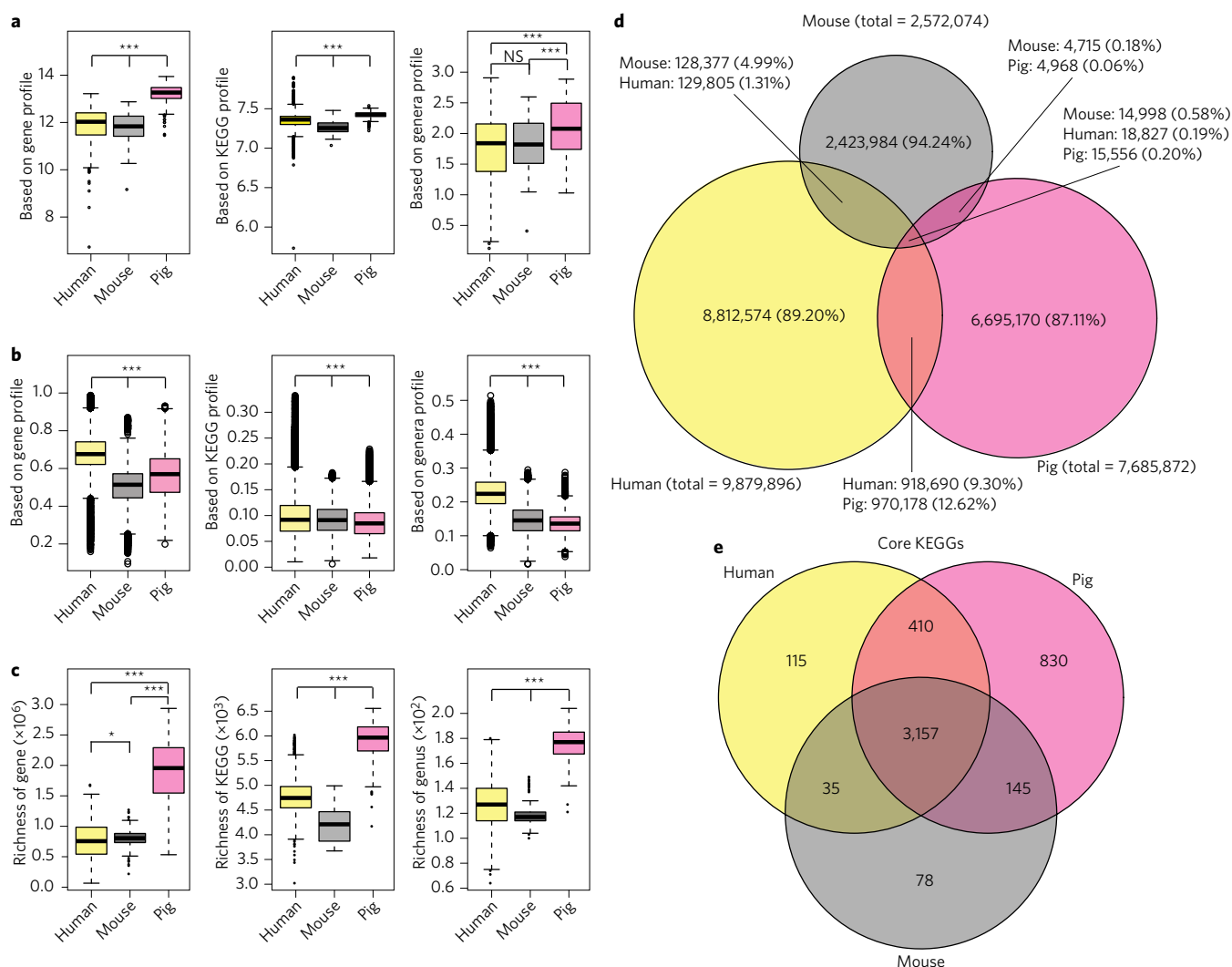
The pig gut metagenome catalogue was compared to the mouse<sup>12</sup> and the integrated human<sup>4</sup> gut metagenome catalogues. The pig catalogue comprised more predicted genes than the mouse catalogue (184 samples), but fewer predicted genes compared to the integrated human gene catalogue (>1,200 samples) (Fig. 1a). The pig metagenome exhibited a higher alpha diversity (Fig. 2a) than both the human and mouse microbiomes, and lower beta diversity than the human metagenome at the gene, genus and KO levels (Fig. 2b). Gene richness, genus richness and KO richness in pigs were higher than those observed in humans and mice (Fig. 2c). Based on a 90% inter-individual sharing within each animal species, a remarkably low percentage of the genes identified in the pig, human and mouse catalogues was shared by all three species (0.20% for pig, 0.19% for human and 0.58% for mouse). The pairwise overlap at the gene level was also modest (pig versus human, 985,734 genes; pig versus mouse 20,524 genes; and mouse versus human 143,375 genes) (Fig. 2d), but was substantially higher for pigs and humans (9.49%) than for mice and humans (1.50%). At the functional level, a far larger proportion of KO functions was shared (3,157 KO identifiers, 66.2% of the total number) (Fig. 2e), representing a functional core in the three mammals (Supplementary Fig. 4). Among these were prominent metabolic functions (carbohydrates, amino acids, cofactors and vitamins, energy metabolism), information processing (membrane transport) and genetic information processing (replication and repair, translation and transcription) that largely correspond to house-keeping functions. Of the KO functional pathways identified in the pig gut metagenome, 78% are present in the human gut metagenome, whereas 96% of the KOs found in the human gut metagenome are present in the pig gut metagenome based on the taxonomically annotated NR genes, supporting the use of the pig as a model for functional studies on the role of the microbiome in health and disease.

To investigate the possible impact of host genetics on the gut microbiome, we selected Chinese female pigs representing six different

breeds, but kept on the same farm and fed the same feed (Fig. 3a and Supplementary Fig. 5). Using non-metric multidimensional scaling (NMDS), the samples separated into three distinct groups at the gene (Fig. 3a) and MGS (Supplementary Fig. 5d) levels. One group comprised the commercial breeds Large White (LW), Binary mixed (HybCN1) and Tertiary mixed (HybCN2), and a second group comprised the native Chinese Bama minipigs and the BaRing pigs. Tibetan pigs, adapted to life at high altitudes<sup>13</sup> separated into a third distinct group (except for one outlier). At the KO level, the Tibetan pigs also clustered separately (Supplementary Fig. 5e). In the data presented in this letter, the maternal and breed effects are partly linked, because the animals from different breeds were not randomly mixed. However, our results suggest an influence of host genetics in shaping the composition of the gut microbiota<sup>14</sup>. More large-scale studies are required to confirm this notion.

To investigate the possible role of age, we used samples from the French and Danish pig sets sampled from 55 to 239 days of age. The analyses revealed an NMDS ordination according to age at the gene level for the French (Fig. 3b) and Danish (Supplementary Fig. 6a) pigs. A similar ordination was found at the KO level in both the French (Supplementary Fig. 6b) and Danish (Supplementary Fig. 6c) pigs.

Danish male, female and castrated male pigs were used to evaluate how sex hormones influence the composition of the gut microbiota. NMDS showed a robust separation between co-housed male and female pigs (Fig. 3c), but not between co-housed castrated males and females (Fig. 3d). Consistently, very few differences were observed comparing castrated males with females (Supplementary Tables 4 and 5). By comparing males and females, 25 genera, 41 species, 75 MGS and 498 KOs exhibited different relative abundances (Supplementary Tables 6–8). At the KO functional level, the most highly represented pathways corresponded to ABC transporters, the ribosome and the phosphotransferase system (Supplementary Table 9). By contrast, only one KO function, representing polyketide synthase PksJ and including the gene encoding polyketide synthase 1/15, distinguished castrated males from females (Supplementary Table 5).

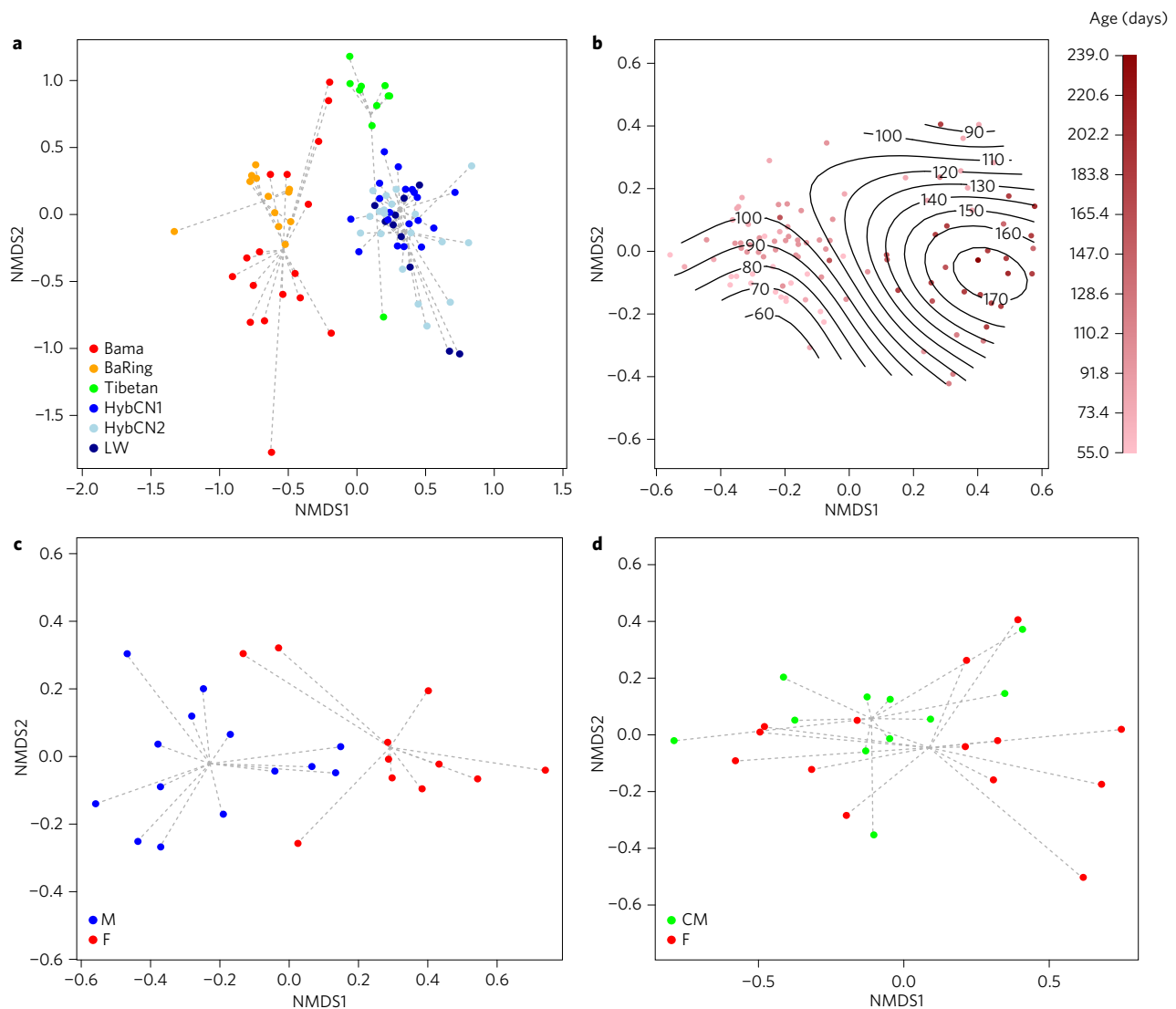


**Figure 2 | Comparison of the gut microbiome gene catalogues of human, mouse and pig.** **a-c**, Alpha diversity (Shannon index) (**a**), beta diversity (Whittaker index) (**b**) and richness (**c**) at the NR gene counts, KO and genera levels. Boxes denote the interquartile range (IQR) between the first and third quartiles (25th and 75th percentiles, respectively) and the line inside denotes the median. Whiskers denote the lowest and highest values within 1.5 times the IQR from the first and third quartiles, respectively. Circles denote outliers beyond the whiskers. **d**, Venn diagram of shared NR genes between the human (yellow), mouse (grey) and pig (pink) catalogues. Only 0.20% of the genes from the pig catalogue are present in the human and mouse catalogues, and the pig catalogue shares 12.82% of the genes with the human catalogue. **e**, Venn diagram of KO functions present in 90% of the individuals in each data set and shared by the human, mouse and pig catalogues. Based on taxonomically annotated NR genes assigned a KEGG function, 3,157 KEGG functions are shared by the pig, human and mouse catalogues, representing 69.3, 84.9 and 92.4% of the total KEGG functions of each catalogue, respectively. These data highlight the consistency of the gut microbiota at the functional level in these three mammalian species and also suggest a significant set of pig-specific KEGG functions.

We next quantified the relative abundances of antibiotic resistance genes (ARGs) in faecal DNA. The NMDS analysis separated the 287 pigs into two groups (Fig. 4a). The first included all Chinese pigs that had been continuously fed low doses of antibiotics and consistently harboured higher abundances of ARGs (Fig. 4b). The second group comprised Danish and French pigs (Supplementary Table 1) that were not fed antibiotics prophylactically. The separation of the Chinese pigs was maintained using total NR gene, MGS or KO function counts, suggesting clear effects of country-specific farm systems and antibiotic supplementation (Supplementary Fig. 7). Administration of antibiotics in the feed was associated with a significant decrease in microbiota richness in the Chinese pigs (Fig. 1a), as predicted<sup>15</sup>. Although both countries and farm systems were identified as confounding factors, several different ARGs are probably involved in the separation of the Chinese pigs from the French and Danish pigs. Indeed vancomycin and teicoplanin ARGs were found to be underrepresented in the Chinese pigs compared to French and

Danish pigs (Fig. 4b). The highest shared prevalence among all pigs was identified to comprise genes encoding resistance to bacitracin, cephalosporin, macrolide, streptogramin B and tetracycline (Fig. 4b). Other ARGs such as genes encoding resistance to chloramphenicol, gentamycin B, kanamycin and neomycin were also present in all pigs but with a much lower abundance in the French and Danish pigs compared to the Chinese pigs (Fig. 4b). In line with the present limited use of fluoroquinolones in European farms, fluoroquinolone resistance genes were not found in the microbiomes of French and Danish pigs.

An increase in the abundance of microbial functional genes related to energy production has been reported in antibiotics-fed pigs<sup>16</sup>. As major antibiotic classes (beta-lactams, aminoglycosides and quinolones) promote the generation of lethal hydroxyl radicals in both gram-negative and -positive bacteria due to increased tricarboxylic acid (TCA) cycle activity<sup>17</sup>, we quantified the abundance of bacterial TCA enzymes and compared those with the relative abundances of ARGs (Supplementary Fig. 8). The microbiomes of Chinese pigs



**Figure 3 | Influence of breed, age and sex on the pig gut microbiota composition, as represented by NMDS plots at the NR gene count level. a**, Effect of breed, shown for the subset of Chinese pigs that included six breeds kept in the same farm and fed the same feed. **b**, Influence of age, shown for the subset of French pigs that ranged from 55 to 239 days old. **c,d**, Effect of sex, shown for the Danish pig subset that comprised either males (M, blue) and females (F, red) fed the same flour-based wet-feed diet co-housed in the Stærsmind farm (**c**) or females (F, red) and castrated males (CM, green) fed the same flour-based dry-feed diet co-housed in the Svindinge farm (**d**).

showed a higher abundance of three TCA enzymes (that is, pyruvate dehydrogenase, succinyl-CoA synthetase and pyruvate carboxylase) compared to French and Danish pigs, resulting in a significant correlation ( $r = 0.95$ ,  $P < 0.001$ ) between the abundance of genes conferring resistance to beta-lactams and the abundance of the gene encoding the pyruvate dehydrogenase E1 component.

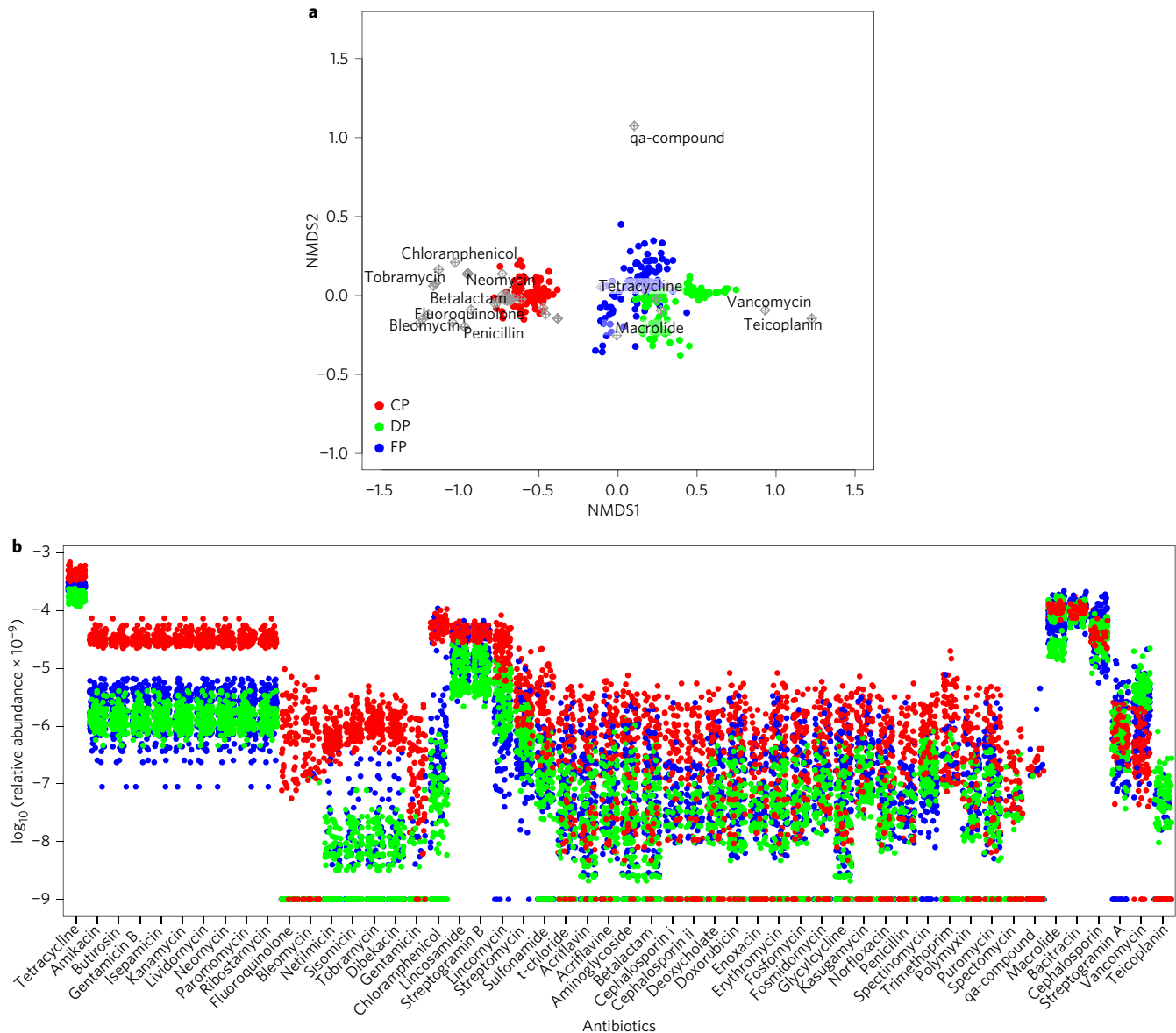
## Discussion

The comprehensive gene catalogue reported here comprising 7,685,872 NR genes will serve as a reference resource for future metagenomics-based studies. The comparison with human and mouse catalogues demonstrated that the overlap between the human and pig microbiome is modest at the gene level but high at the KO function level. Still, our results point to a closer similarity between pig and human gut microbiomes than between mouse and human gut microbiomes. We show how sex, age and host genetics are likely to influence the gut microbial communities and metabolic profiles.

We report a common set of ARGs found in all pigs, regardless of the country of origin or supplementation with antibiotics. The

use of antibiotics as growth promotants has been banned in Europe since 2006. It is noteworthy that our results demonstrate a higher load of ARGs in the Chinese pigs that are continuously fed antibiotics than in the Danish and French pigs, strongly indicating that elimination of antibiotics from animal feed is an efficient means to reduce the risk of ARG dissemination in relation to pig production. However, French and Danish pigs retained a wide range of ARGs, providing a picture of ARG reservoirs in non-antibiotic-fed pigs in France and Denmark. Globally, the significant ARG load in all pigs reflects the long-term persistence of ARGs, irrespective of current antibiotics use, and monitoring the prevalence of ARGs through the continuum of ecosystems, in line with the One Health Initiative, will be a valuable strategy to evaluate ARG loads in farm animals and pave the way for more sustainable farm systems<sup>11</sup>.

Combined with the recently published pig genome, the pig gut gene catalogue will accelerate research that aims at deciphering the complex interactions between microbiota and hosts. Integration of phenotypic, genomic and metagenomic data will provide key biological information



**Figure 4 | Prevalence of antibiotic resistance genes (ARGs).** **a**, NMDS biplot of the ARGs reveals a clear separation of Chinese pigs compared to French and Danish pigs. Grey diamonds represent the individual ARG ordination onto the two-dimensional space, with the names of a subset of ARGs given in black. **b**, Relative abundances ( $\log_{10}$  scale) of ARGs found in each individual. ARGs are ordered on the x axis and each dot represents a Chinese (red), French (blue) or Danish (green) pig. The higher the vertical position of the dots on the y axis, the higher the relative abundance of the ARGs. CP, Chinese pigs; DP, Danish pigs; FP, French pigs.

for future biomedical research, as well as in translational research towards more sustainable knowledge-based pig farming.

## Methods

**Faecal sample collection and DNA extraction.** Fresh faeces were sampled at the rectum in 287 pigs of various ages and breeds, on different diets in French, Danish and Chinese farms (Supplementary Table 1), immediately frozen in liquid nitrogen or dry ice and stored at  $-80^{\circ}\text{C}$ . DNA extractions were performed using 200 mg faeces per sample following the method reported by Qin *et al.*<sup>3</sup> except that cell lysis was performed by bead beating the samples twice for 30 s, with an incubation of 2 min on ice between beatings. The resulting concentrations of faecal DNA were measured by Nanodrop.

**DNA sequencing.** Following the manufacturer's instructions (Illumina), we constructed one DNA paired-end (PE) library with an insert size of 350 base pairs (bp) for each sample. All samples were sequenced by an Illumina GAIIx and HiSeq 2000 and 100 bp PE strategy<sup>3</sup>.

**DNA assembly and construction of the gene catalogue.** Raw reads were filtered to remove adaptor contamination, low-quality reads with 'N' bases and host genomic DNA (NCBI accession no. NC\_010443). An average of 2.1% of the raw reads

corresponding to host (pig) genome DNA were removed. The remaining reads were referred to as high-quality reads. In total, we obtained 1,758 Gb high-quality data with an average of 6.13 Gb per sample. To construct a comprehensive pig gut microbial gene catalogue, we assembled the Illumina reads from each sample into longer contigs with the SOAPdenovo software (v1.06)<sup>18</sup>. A total of 73.7% of the reads were assembled into 32 million contigs with a length exceeding 500 bases, giving a total contig length of 45.7 Gb. To avoid errors generated during the assembly process, all the short reads were mapped back to the assembled contigs, whereupon single bases and indels were corrected according to the mapping depth. GeneMark (v2.7)<sup>19</sup> was used to predict open reading frames (ORFs) in contigs obtained for each sample. A NR gene set was constructed by pairwise comparison of all genes in all samples, using BLAT (ref. 20) and the criteria of identity of  $>95\%$  and overlap of  $>90\%$ . An overall shorter average ORF length and N50 assembled contig length (680 bases and 1.87 kb, respectively) was obtained for the pig gut catalogue compared to both human (721 bases and 5.0 kb) and mouse (762 bases and 6.6 kb) counterparts. Taxonomic assignments were made with CARMA3 (ref. 21) on the basis of BLASTP (ref. 22) against the NCBI-NR database. Functional annotations were made by BLASTP search against eggnoG (ref. 23) and KEGG v48.2 (ref. 24) databases.

Abundance profiles for the catalogue genes were co-abundance clustered using the canopy-clustering algorithm<sup>25</sup>. The combined results of two clustering rounds, one using 75% quantile profiles and the Pearson correlation coefficient and one using

mean profiles and the Spearman correlation coefficient, were joined into a single set of MGS using 500 genes as the lower cutoff. The MGS set was checked and reduced for redundancy and filtered using a Spearman gene inclusion filter.

**Ordination analysis.** Different methods for ordination analysis implemented in the Vegan R package<sup>26</sup> were employed at the gene, MGS, KEGG, taxonomic (genus, phylum, species) and ARG levels. Dissimilarity between pairs of samples was defined using the Bray–Curtis dissimilarity index<sup>27</sup> and visualized using two-dimensional NMDS<sup>28</sup>.

**Differential abundance analysis.** Differential abundance analysis of the metagenomic data sets was performed based on a zero-inflated Gaussian mixture model implemented in the fitZig function of the metagenomeSeq R package<sup>29</sup>. Different models were implemented to correct for confounding covariates according to the specific data set as follow:

$$y_{ijm} = \text{Country}_i + \text{Sex}_j + e_{ijm}$$

where  $y_{ijm}$  is the  $l$ th individual normalized record, country (three categories) and sex (three categories) are fixed effects and  $e_{ijm}$  is the residual. This model was fitted taking into consideration the common environmental information (country and sex) for the whole data set ( $n = 287$ ) and was applied to identified differential abundant MGS/KEGG and ARGs across countries. In all scenarios, correction for multiple testing was performed using the  $q$  value, and the cutoff of the differential abundance was set at  $q \leq 0.05$ . Correlation between the abundance of ARGs and enzymes from the TCA cycle were determined throughout the entire data set ( $n = 287$ ) and also by countries using the Spearman's rank correlation coefficient in R.

To identify the sex difference in the pig gut metagenome, we focused the analysis on two subsets of animals from the Danish samples. The first comparison included 25 pigs (11 males and 14 females), all from the same farm (Stærsminde), the same breed (Landrace  $\times$  Yorkshire)  $\times$  Duroc and fed the same diet (Supplementary Table 1). In the second farm (Svinginge) comparisons of 20 pigs (10 females and 10 castrated males) from the same breed (Landrace  $\times$  Yorkshire)  $\times$  Duroc and on the same diet were included. Differential abundant genes or enzymes identified between sex were mapped on the KEGG metabolic pathways using the iPath2.0 tool<sup>30</sup>.

**Data availability.** Gut metagenome sequences have been deposited in the European Nucleotide Archive (ENA) under accession code PRJEB11755. The gene sequences and related profiles at gene and KO levels are released at GiGaDB (<http://gigadb.org/dataset/view/id/100187/token/F4CDHYruxobOKmSE>).

Received 26 February 2016; accepted 2 August 2016;  
published 19 September 2016

## References

- Lunny, J. K. Advances in swine biomedical model genomics. *Int. J. Biol. Sci.* **3**, 179–184 (2007).
- Sommer, F. & Bäckhed, F. The gut microbiota—masters of host development and physiology. *Nat. Rev. Microbiol.* **11**, 227–238 (2013).
- Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
- Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
- Groenen, M. A. M. *et al.* Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398 (2012).
- Osei Sekyere, J. & Osei Sekyere, J. Antibiotic types and handling practices in disease management among pig farms in Ashanti region, Ghana. *J. Vet. Med.* **2014**, e531952 (2014).
- Aubry-Damon, H. *et al.* Antimicrobial resistance in commensal flora of pig farmers. *Emerg. Infect. Dis.* **10**, 873–879 (2004).
- Sayah, R. S., Kaneene, J. B., Johnson, Y. & Miller, R. Patterns of antimicrobial resistance observed in *Escherichia coli* isolates obtained from domestic- and wild-animal faecal samples, human septage, and surface water. *Appl. Environ. Microbiol.* **71**, 1394–1404 (2005).
- Lammie, S. L. & Hughes, J. M. Antimicrobial resistance, food safety, and one health: the need for convergence. *Annu. Rev. Food Sci. Technol.* **7**, 287–312 (2016).
- Sanderson, H., Fricker, C., Brown, R. S., Majury, A. & Liss, S. N. Antibiotic resistance genes as an emerging environmental contaminant. *Environ. Rev.* **24**, 205–218 (2016).
- Williams-Nguyen, J. *et al.* Antibiotics and antibiotic resistance in agroecosystems: state of the science. *J. Environ. Qual.* **45**, 394–406 (2016).
- Xiao, L. *et al.* A catalog of the mouse gut metagenome. *Nat. Biotechnol.* **33**, 1103–1108 (2015).
- Ai, H. *et al.* Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics* **15**, 834 (2014).
- 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. USA* **107**, 18933–18938 (2010).
- Modi, S. R., Collins, J. J. & Relman, D. A. Antibiotics and the gut microbiota. *J. Clin. Invest.* **124**, 4212–4218 (2014).
- Loof, T. *et al.* In-feed antibiotic effects on the swine intestinal microbiome. *Proc. Natl Acad. Sci. USA* **109**, 1691–1696 (2012).
- Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* **130**, 797–810 (2007).
- Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**, 18 (2012).
- Besemer, J., Lomsadze, A. & Borodovsky, M. Genemarks: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **29**, 2607–2618 (2001).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
- Gerlach, W. & Stoye, J. Taxonomic classification of metagenomic shotgun sequences with CARMA3. *Nucleic Acids Res.* **39**, e91 (2011).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Powell, S. *et al.* eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* **42**, D231–D239 (2014).
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2015).
- Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
- Dixon, P. & Palmer, M. W. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
- Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
- Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* **27**, 125–140 (1962).
- Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **10**, 1200–1202 (2013).
- Yamada, T., Letunic, I., Okuda, S., Kanehisa, M. & Bork, P. iPath2.0: interactive pathway explorer. *Nucleic Acids Res.* **39**, W412–W415 (2011).

## Acknowledgements

This work was supported financially by the INRA metaprogramme Meta-omics of Microbial Ecosystems (MEM), the Animal Genetics Division of INRA, the Biology Department of the University of Copenhagen and BGI-Shenzhen. Y.R.C. was funded by the European Union, in the framework of the Marie-Curie FP7 COFUND People Program, through the award of an AgreeSkills' fellowship (grant agreement no. 267196) linked to the METALIT project funded by the INRA MEM metaprogramme. This study was supported by the Shenzhen Municipal Government of China (grant nos. JCYJ20140418095735538, DRC-SZ[2015]162, CXB201108250098A and CXZZ20150330171521403), the National Transgenic Project of China (2014ZX0801007B), the Science and Technology Plan of Shenzhen City, the Shenzhen-Hong Kong Innovation Circle Research Fund, China (SGLH20120928140607107). The work was also supported by the Danish Innovation Fund (grant nos. 4105-00011A and 0603-00774B) and Metagenopolis grant ANR-11-DPBS-0001. The authors acknowledge technical support for faecal stool sampling provided by the INRA experimental units (Y. Billon, UE GENESI at le Magneraud; C. Jaeger and H. Demay, UMR SENAH at Saint Gilles; N. Müller, UE0787 at le Rheu; E. Venturi and his team, UE PAO at Nouzilly; J.-L. Gourdière and D. Renaudeau, URZ in the Guadeloupe; J.-J. Leplat, UMR GABI at Jouy-en-Josas). The authors thank K. Pedersen at Svinginge farm, T. Povlsen at Blaesborg farm, V. Jensen and K. Jensen at Stærsminde farm and H. Loft Hansen for technical assistance in relation to collection of faecal samples from the Danish farms. The authors also thank X. Liu and Zhigang at the Institute of Allergy & Immunology, Shenzhen University, and L. Fang Liang, Q. Wei and Y. Li at the technical unit of the Chinese farms for help with faecal sampling and administration. The authors thank C. Denis (INRA, UMR GABI, Jouy-en-Josas) and F. Levenez (INRA, UMR MICALIS and MetaGenoPolis, Jouy-en-Josas) for faecal DNA preparation. The authors acknowledge the @BRIDGE platform (INRA, Jouy-en-Josas) for safe storage of samples. The authors are grateful to J. Nathaniel Paulson for kind assistance with the use of the R package referred to as metagenomeSeq, and J. Lunney (Animal Parasitic Diseases Laboratory, ARS, USDA, MD, Beltsville, USA) for discussions and the contribution to manuscript editing.

## Author contributions

C.R.-G., J.W., L.M., E.M., J.D., S.D.E. and K.K. conceived and designed the project. K.K., C.R.-G., L.M., S.D.E., J.W., L.X., J.E., P.K., Y.R.-C., X.X., X.L., N.J.K. and Z.X. monitored the project. A.Ø.P., C.L. and E.P. collected samples and performed experiments. K.K., C.R.-G., L.M., J.W., J.E., Y.R.-C., P.K., N.P., E.L.C., E.P., S.D.E., Z.X., Q.F., S.L., A.Ø.P. and J.L. analysed and interpreted the data. K.K., C.R.-G., J.E., Y.R.-C., S.D.E., L.M., P.K., H.J. and L.X. wrote the paper. All authors commented on the manuscript.

## Additional information

Supplementary information is available for this paper. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.K., C.R.-G. and J.W.

## Competing interests

The authors declare no competing financial interests.