Tittle:

A pilot RNA-seq study in 40 Pietrain ejaculates to characterize the porcine sperm microbiome

Marta Gòdia[a], Yuliaxis Ramayo-Caldas[b,c], Laura M. Zingaretti[a], Laila Darwich[d,e], Samantha López[f], Joan E. Rodriguez-Gil[g], Marc Yeste[h], Armand Sánchez[a,i], Àlex Clop[a,j]*

a. Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, 08193, Cerdanyola del Vallès, Catalonia, Spain

b. Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, 08140, Caldes de Montbui, Catalonia, Spain.

c. GABI, INRA, AgroParisTech, Université Paris-Saclay, UMR 1313, Jouy-en-Josas, France.

d. Departament de Sanitat i d'Anatomia Animals, Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain.

e. IRTA, Centre de Recerca en Sanitat Animal (CReSA, IRTA-UAB), Campus de la Universitat Autònoma de Barcelona, 08193 Bellaterra, Catalonia, Spain.

f. Department of condensed matter physics, Faculty of Physics 08028, Universitat de Barcelona, Barcelona, Cataloina, Spain

g. Department of Animal Medicine and Surgery, School of Veterinary Sciences, Universitat Autonoma de Barcelona, 08193, Cerdanyola del Vallès, Catalonia, Spain

h. Biotechnology of Animal and Human Reproduction (TechnoSperm), Unit of Cell Biology, Department of Biology, Institute of Food and Agricultural Technology, Faculty of Sciences, University of Girona, E-17003 Girona, Catalonia, Spain.

i. Departament de Ciència Animal i dels Aliments, School of Veterinary Sciences, Universitat Autonoma de Barcelona, 08193, Cerdanyola del Vallès, Catalonia, Spain

j. Consejo Superior de Investigaciones Científicas (CSIC), 08003, Barcelona, Catalonia, Spain

E-mail adresses:

M. Gòdia: marta.godia@cragenomica.es

Y. Ramayo-Caldas: yuliaxis.ramayo@irta.cat

L. M. Zingaretti: laura.zingaretti@cragenomica.es

L. Darwich: Laila.darwich@uab.cat

S. López: samantha.lopezm94@gmail.com

J. E. Rodriguez-Gil: JuanEnrique.Rodriguez@uab.cat

M. Yeste: marc.yeste@udg.edu

A. Sánchez: armand.sanchez@uab.cat

A. Clop: alex.clop@cragenomica.es

*Corresponding author:

alex.clop@cragenomica.es; Tel. +34 935636600 ext 3353

Abstract

The microbiome plays a key role in homeostasis and health and it has been also linked to fertility and semen quality in several animal species including swine. Despite the more than likely importance of sperm bacteria on the boar's reproductive ability and the dissemination of pathogens and antimicrobial resistance genes, the high throughput characterization of the swine sperm microbiome remains scarce. We carried RNA-seq on 40 ejaculates each from a different Pietrain boar and found that a proportion of the sequencing reads did not map to the *Sus scrofa* genome. The current study aimed at using these reads not belonging to pig to carry a pilot study to profile the boar sperm bacterial population and its relation with 7 semen quality traits.

We found that the boar sperm contains a broad population of bacteria. The most abundant phyla were *Proteobacteria* (39.1%), *Firmicutes* (27.5%), *Actinobacteria* (14.9%) and *Bacteroidetes* (5.7%). The predominant species contaminated sperm after ejaculation from soil, faeces and water sources (*Bacillus megaterium*, *Brachybacterium faecium*, *Bacillus coagulans*). Some potential pathogens were also found but at relatively low levels (*Escherichia coli*, *Clostridioides difficile*, *Clostridium perfringens*, *Clostridium botulinum* and *Mycobacterium tuberculosis*). We also identified 3 potential antibiotic resistant genes from *E. coli* against chloramphenicol, *Neisseria meningitidis* against spectinomycin and *Staphylococcus aureus* against linezolid. None of these genes were highly abundant. Finally, we classified the ejaculates into categories according to their bacterial features and semen quality parameters and identified two categories that significantly differed for 5 semen quality traits and 13 bacterial features including the genera *Acinetobacter*,

Stenotrophomonas and *Rhodobacter*. Our results show that boar semen contains a bacterial community, including potential pathogens and putative antibiotic resistance genes, and that these bacteria may affect its reproductive performance.

Keywords

Pig, spermatozoa, microbiome, RNA-seq, sperm quality

1. Introduction

Scientific research has led to the discovery that many compartments of the animal organism contain a rich and complex population of microorganisms known as microbiota, which plays a crucial role in physiological homeostasis and health [1-3] including sperm quality and male fertility [4, 5]. The male's reproductive ability is represented by a set of traits that are important for human health and for the efficiency and sustainability of animal production. In swine, semen quality is regularly measured in the artificial insemination studs as a proxy of the fertilization ability of that sample. Growing research is being devoted to understanding the biological basis and identifying molecular markers linked to semen quality in humans and other animal species. As the presence of bacterial communities in ejaculates is common and the microbiome is popping up as a big contributor of a broad range of phenotypes, several studies have been carried in the field of men fertility [4, 6, 7] and boar sperm quality [8, 9]. Weng et al. [4] identified a complex population of bacteria in human sperm but most interestingly, found that the abundance of some bacteria was related to male fertility. *Lactobacillus crispatus*, *Gardnerella vaginalis* and *Lactobacillus acidophilus* were more abundant in the fertile samples whilst *Prevotella vibia* and *Haemophilus parainfluenzae* were present at higher proportion in the unfertile sperm [4]. In a more recent study, a group led by Stephen Krawetz [10] used sperm RNA-seq datasets to identify transcripts of bacterial origin and shed light to the bacterial composition of an ejaculate. They found a diverse bacterial population mostly characterized by members of the phyla *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* [10].

In pigs, the presence of bacteria in sperm is well documented and bacterial populations in ejaculates are common [11-13]. In pigs, most of the bacteria present in semen ejaculates have an external origin and have contaminated the sperm after ejaculation. The most abundant sources of contaminations are the prepucial diverticulum and hair [11], the sinks and drains of the stud, the utensils used for ejaculate collection and transfer as well as the laboratory surfaces where the ejaculates are being processed [14]. The presence of bacteria in sperm is of further concern within the One Health concept as commercial sperm doses in the livestock industry can be a major contributor on the dissemination of bacterial pathogens and antibiotic resistance genes (ARGs) [15]. Ubeda and co-authors, using cell culture, concluded that the most abundant bacteria in pig semen were from the Enterobacteriaceae family and included, in order of abundance, S*erratia marcescens*, *Klebsiella oxytoca*, *Providencia stuartii*, *Morganella morganii*, *Proteus mirabilis*, and *Escherichia coli*. *S. marcescens*, *K. oxytoca*, *M. morganii*, or *P. mirabilis* were negatively associated with sperm quality [8]. Schulze also recently identified the presence of several species of *Lactobacillus* and an association, *in vitro*, between the abundance of *Lactobacillus buchneri* and sperm motility, mitochondrial activity and membrane integrity and *Lactobacillus animalis* with motility [16]. To control bacterial growth in sperm, antimicrobials are commonly added to semen extenders [13]. Nonetheless, bacteria in these extended ejaculates can be still present due to incomplete efficiency of the antibiotics which could be partially caused by the expression of ARGs by these bacteria. Current high throughput sequencing technologies provide unprecedented capacity to study and expose the complexity of microbial ecosystems. Recently, Even et al. explored for first

time the pig sperm microbiome using high throughput sequencing of the 16S bacterial gene. The aim of their study was to identify the factors that influence the sperm microbiome and to assess the adequacy of this technique to routinely monitor the sperm bacterial population [12]. The authors nicely showed that the stud has an effect on the bacterial composition of the porcine semen [12]. Although the experimental design did not allow disentangling in detail the stud related factors that shape the seminal microbiota, the flooring type itself (sawdust or slatted floors) showed association with the microbiome composition and diversity [12]. They also found that diluting the ejaculates with extenders, which contain antibiotics, reduces the bacterial diversity in a sample and also contributes reducing the variability in the bacterial diversity between ejaculates [12]. The aim of our study was to characterize the composition of the boar sperm microbiome exploiting a RNA-seq dataset on extended sperm from 40 pigs and interrogate the existence of a potential link between the sperm microbiome and semen quality traits.

2. Materials and methods

2.1. Sample collection, purification and phenotyping

Specialized professionals obtained fresh ejaculates from 40 Pietrain boars from three different commercial farms located in Catalonia (~42 ºN, ~2 ºE), with the gloved-hand method. The farms contained between 114 and 140 boars in 6 squared meter pens each harboring 6 boars. All farms had sawdust flooring, did not use air filtration system and employed similar semen collection and processing practices. Ejaculates were collected between March 2015 and January 2017 and boar ages ranged from 9 to 55 months old. After collection,

157 the ejaculates were immediately diluted (1:1) with Androstar® Plus and kept at
158 26 °C for up to 6 hours until they were further diluted (1:2) in Androstar® Plus.
159 Androstar® Plus is a boar semen extender that contains the following
160 antibiotics: apramycin sulphate, cephalosporin – third generation - and
161 gentamicin sulphate. The extended samples were then kept at 16 °C for 6-10
162 additional h until they were processed in our laboratory for phenotyping and
163 spermatozoa purification. Seven sperm phenotypes were measured in the 40
164 samples as previously described by Godia et al. [17]. Phenotypes included the
165 percentage of viable sperm cells after 5 min of incubation at 37 ºC (VIAB_5),
166 the percentage of viable sperm cells after 90 min incubation at 37 ºC (VIAB_90),
167 percentage of cells with abnormal acrosomes after the 5 min (ACRO_5) and
168 the 90 min (ACRO_90) incubation, the percentage of motile cells after 5 min
169 (MT_5) and 90 min (MT_90) incubation and the percentage of membrane
170 functional spermatozoa after an osmotic stress (ORT, Osmotic Resistance
171 Test). VIAB_5, VIAB_90, ACRO_5 and ACRO_90 were measured by staining
172 the samples with the eosin-negrosin technique after 5 and 90 min incubation at
173 37 °C following the protocol described by Bamba [18]. MT_5 and MT_90 were
174 measured with the computer-assisted semen analysis (CASA) system
175 (Integrated Sperm Analysis System V1.0; Proiser). To calculate ORT the
176 spermatozoa were incubated at 37 °C for 10 min on iso- and hypo-osmotic
177 solutions using the method described by Rodríguez-Gil and Rigau [19].
178 Normal motile spermatozoa were subsequently purified using the BoviPure™
179 colloidal silica particles reagent (Nidacon; Mölndal, Sweden) as detailed by
180 Gòdia et al. [17]. Briefly, the volume of sperm that was used varied according
181 to the sperm concentration, with a maximum of 1 billion cells and not exceeding

11 mL. The manufacturer's recommendation of a minimum volume ratio of 25% diluted BoviPure™ / semen was maintained. After centrifugation following the manufacturer's protocol, the cell pellet was washed once with RNase-free phosphate buffer saline (PBS) and then resuspended in 1 mL of RNase-free PBS for optical inspection to confirm the removal of somatic cells. For all samples, aliquots containing ~40 million spermatozoa were then centrifuged and the resulting pellet was stored at -80 °C in 1 mL of Trizol® until further processed for RNA extraction.

2.2. RNA extraction, qPCR validation, library prep, sequencing

RNA was extracted from sperm pellets using a standard Trizol® approach and treated with TURBO DNA-free™ Kit (Invitrogen; Carlsbad, USA) [17]. RNA samples were subjected to RT-qPCR assays to validate the presence of spermatozoa RNA with primers targeting the *PRM1* gene, the absence of RNA from contaminating diploid cells (mainly leukocytes and keratinocytes) using primers against the somatic gene *PTPRC* and the absence of genomic DNA using primers targeting an intergenic region [17]. Total RNA was subjected to mammalian ribosomal RNA (rRNA) depletion with the Ribo-Zero Gold rRNA Removal Kit (Illumina, CA, USA). RNA-seq libraries were prepared with SMARTer Universal Low Input RNA library Prep kit (Clontech, France) and sequenced in an Illumina's HiSeq2000/2500 system to generate 75 base pair long paired end reads. These RNA-seq datasets were initially analyzed to characterize the boar sperm transcriptome [20] and circular RNAome [21]. The RNA-seq data used in this study (total RNA-seq runs) is accessible at the NCBI's under the SRA study accession SRP183646.

2.3. Bioinformatics and statistical analysis

2.3.1. Identification of RNA molecules of bacterial origin

RNA-seq reads of low quality and adaptor contaminations were removed with Trimmomatic v.0.36 [22]. Filtered reads were then mapped to the *Sus scrofa* genome (Sscrofa11.1) with HISAT2 v.2.1.0 [23] with default parameters except "--max seeds 30" and "-k 2". The reads that did not map to Sscrofa11.1 were screened against the catalogue of porcine Transposable Elements from the Repbase database [24] with HISAT2 v.2.1.0 [23].

The reads that remained unmapped were taxonomically classified and quantified with Kraken v.0.10.5 [25] with a threshold score of 0.15 and using the default database that includes NCBI taxonomic information and complete genomes from RefSeq of archaeal, bacteria, phage and viral domains. Only the bacterial-assigned reads were kept for further analysis. The number of reads assigned to a given taxon was normalized by sequencing depth, as counts per million (CPM).

For ease of readability, we refer to each bacterial taxon identified in the study, from phyla to species, as bacterial feature. The list of potential pathogens in swine was extracted from the Professional Pig Community pig333 site (www.pig333.com/pig-diseases), The Pig Site (https://thepigsite.com/disease-and-welfare/managing-disease/bacteria) and The Swine Health Information Center (https://www.swinehealth.org/swine-bacterial-disease-matrix/). The list of bacterial agents and diseases in each of these sources is available at Table S1.

2.3.2. Detection of antimicrobial resistance genes

Unmapped reads were also subjected to identification and relative abundance quantification of ARGs. ARGs were identified using BLASTN v.2.7.1 [26] with

100% percentage identity using the Comprehensive Antibiotic Resistance Database (CARD) v.3.0.0 [27]. The number of reads for each ARG was normalized by sequencing depth, as CPM. The read coverage across ARGs of point mutations was individually visualized using R v.3.5.3 [28].

The relationship between the abundance of each ARG and the abundance of their corresponding bacteria was calculated with “lm” function from R [28]. The adjusted R-squared was provided by the lm function and corresponds to the Wherry’s formula [29].

2.3.3. Relation between bacterial abundance and semen quality traits

The raw phenotypes were corrected by environmental factors including farm of origin, age of boar when sampled and season-year of sampling using a standard linear model [28]. We assessed the relationship between the semen quality traits and the bacterial features of the sperm microbiome using LINK-HD, an integrative methodology designed to deal with the compositional nature of microbial datasets [30]. The methodology is an extension of Principal Component Analysis (PCA) suitable for analyzing several sources of data that share a common set of observations. It outputs a matrix that is often referred to as the compromise. An eigen-analysis of this matrix not only allows a graphical representation of the samples in a plane, but also allows them to be grouped using a standard cluster methodology like k-means. This analysis included the 7 corrected phenotypes and the bacterial features (N=733) with average CPM ≥ 1 and representing more than 0.001% of the total bacterial read counts. We use the cluster classification derived from the compromise structure to perform variable selection through the fitZig function from the metagenomeSeq package v.1.28.2 [31]. fitZig implements an expectation- maximization algorithm (EM) to

estimate the differential abundance of taxa using a Zero-Inflated Gaussian (Zig) distribution that takes their sparse nature into account.

3. Results

3.1. RNA-seq statistics

We carried RNA-seq on 40 extended ejaculates each from a different Pietrain pig and obtained an average of 40.7 million reads per sample. In average, 98.5% of the reads passed the quality control and 82.7% mapped to the porcine genome (Sscrofa11.1). A tiny proportion (0.012%) of the unmapped reads aligned to Repbase [24] and 25.1% (an average of 1.7 million reads per library) mapped to microbial genomes with Kraken (Table S2).

3.2. Description of the boar sperm microbiome

We identified 733 bacterial features with average abundance ≥ 1 CPM and representing more than 0.001% of the total bacterial read counts. The total bacterial abundance across samples varied between 2,241 and 180,624 CPMs (Fig. 1 and Table S3). The average and median abundances of bacterial reads were 20,149 and 9,785 CPM, respectively and 3 ejaculates had more than 70,000 bacterial CPM (Fig. 1). The bacterial features included 15 phyla (Table S3). The most abundant phyla were *Proteobacteria*, with an average of 39.1% of bacterial reads, *Firmicutes* (27.5%), *Actinobacteria* (14.9%) and *Bacteroidetes* (5.7%) (Fig. 2 and Table S3). At the species level, the analysis identified 254 bacterial species (Table S3). The most abundant species were, in this order, *Bacillus megaterium* (868 CPMs and 4.3% of the bacterial reads), *Brachybacterium faecium* (3.3%), *Bacillus coagulans* (1.2%) and *Campylobacter hominis* (1.0%) (Table 1).

3.3. Boar sperm safety: pathogens and antibiotic resistance genes

We found 12 potentially pathogenic species of bacteria with average abundance ≥ 1 CPM and representing more than 0.001% of the total bacterial read counts but only 7 displayed CPM > 5. These were, in this order: *Escherichia coli*, *Clostridioides difficile*, *Clostridium perfringens*, *Clostridium botulinum*, *Mycobacterium tuberculosis*, *Mycoplasma hyopneumoniae* and *Campylobacter jejuni* (Table 2). With the exception of *E. coli* and *C. difficile*, which ranked 8th and 22nd in the list of most abundant bacterial species, with 137 and 50 CPM, respectively, these potential bugs were in general displaying low relative abundance in our samples (Table 2). While nearly all the samples contained at least traces of these bacteria, *M. tuberculosis* was only present in 6 samples and it presented moderate abundances (between 28 and 84 CPMs) in all of them (Table S3).

We also searched for ARG with average CPM ≥ 1 and found 3 candidates, including ARO:3003497, *Neisseria meningitidis* 16S rRNA mutation conferring resistance to spectinomycin; ARO:3004058, *Staphylococcus aureus* 23S rRNA with mutation conferring resistance to linezolid and ARO:3004150, *E. coli* 23S rRNA with mutation conferring resistance to chloramphenicol. Moreover, all the samples presented CPM ≥ 1 for these 3 ARGs (Table 2).

3.4. Relationship between the sperm microbiome and semen quality

To identify potential relationships between bacterial abundances and semen quality we employed Link-HD [30], a recently developed tool based on STATIS methodology to integrate heterogeneous datasets. This approach analyzes different types of variables measured on the same samples, here bacterial abundance and semen quality phenotypes. To the end, the tool turns each raw

data into cross-product matrix, computed on the distances between samples, which are then combined in a common configuration named compromise. A classical Principal Component Analysis (PCA) decomposes the compromise variance into orthogonal components and data structure can be easily recovered using standard clustering techniques. In this study, the samples were clustered into categories according to their microbiome and their semen quality. We included the 733 bacterial features (from phyla to species in Table S3) and 7 semen quality traits (Table S4). Link-HD structured the purified ejaculates into 2 clusters with 30 (cluster 1) and 10 (cluster 2) samples each (Fig. 3 and Table S5). The analysis also recovers the contribution of each feature into the common structure, which facilitates the interpretability of the results. We found that the 7 semen traits and 67 of the 733 bacterial features associated with the whole-compromise structure (Table S6).

We then compared the distribution of these 7 phenotypes and 67 bacterial features in each cluster. The 2 categories showed statistically significant differences for 5 traits. MT_5 and MT_90 did not differ between both groups (Table S7). The feature abundances between the 2 clusters were compared. Thirteen bacterial features resulted in nominal significant differences between clusters (Table 3). These included the genus *Acinetobacter*, *Stenotrophomonas* and *Rhodobacter* (Table 3).

4. Discussion

4.1. Technical considerations

We carried RNA-seq on the extended sperm from 40 pigs with the aim to characterize the boar semen transcriptome in relation to sperm quality. We

hypothesized that a proportion of the sequences that did not map to the pig genome (Sscrofa11.1), between 9 and 31% of the reads (Table S2), could correspond to bacterial transcripts. We identified a rich population of bacteria with a diverse abundance profile between the ejaculates. Despite the fact that the processed extended sperm contained antibiotics and that we treated these samples to remove micro-organisms, we found evidences of bacterial presence in their sequenced RNAs. This indicates that the extender did not eliminate or inactivate all the bacteria present in the ejaculate. We can even hypothesize that these bacteria were viable and transcriptionally active at the time that we processed and froze the samples prior to RNA extraction. Dead bacteria would release their RNA content to the extracellular milieu and this would be degraded by action of the ubiquitous extracellular RNases. However, it also seems that the initial bacterial burden in sperm did not experience an exponential growth during the incubation time (12-16 h) in extended sperm. Bacterial growth follows an exponential pattern with a slope that is dependent on the generational interval [32]. Our measure from the RNA-seq datasets, with total bacterial abundances ranging between 180,000 and 2,241 CPM, and a median of 9,785 CPM, suggest that these bacteria did not proliferate at high rates in our samples possibly due to the effect of the antibiotics. In addition, it is even possible that these antibiotics promote a positive selection for the resistant bacteria. In fact, we observed the presence of 3 ARGs that confer resistance to spectinomycin, linezolid and chloramphenicol. However, all this remains speculative and only classical microbiology tests can ascertain the viability of the cells.

RNA-seq has several particular characteristics when compared to other high throughput evaluations of bacterial communities. First, it allows exploring gene expression and thus assessing the functional activity of the microbiome. For this reason, RNA-seq based quantification is biased towards the identification of the active bacteria. Second, it allows discriminating between active viable and not-viable or dormant microorganisms as the first have active gene expression. Third, it has higher resolution than the analyses targeting exclusively the 16S gene as RNA-seq targets a larger portion of the bacterial genome [33]. However, we used the Kraken metagenomics tool [25] which was designed to quantify the abundance of bacteria based in their DNA. Kraken has been already previously used to characterize the sperm microbiome using RNA-seq datasets in human [10]. While meta-genomics strictly focuses on the abundance of bacterial specimens, meta-transcriptomics informs on the expression of their genes and thus the function and activity of these micro-organisms in the sample. Our data provides a quantification of each bacterium based in the overall expression of their transcripts which accounts for both the bacterial abundance and their gene expression activity and have the additional advantage to account for active microorganisms. In other words, we cannot state without uncertainty whether one bacterium is more abundant than another in one sample but we can assume that this is the most likely scenario as in part, our measures are reflecting these abundances. For this reason and to ease the message provided in this manuscript, we have referred to bacterial abundance throughout the article.

Our experiment is a pilot study based on a small dataset of 40 ejaculates, each from a different boar of the same breed (Pietrain) and representing only 3 studs

with similar management conditions and geographic location on the same climatic zone. Although the information is relevant as little is known on the microbiome composition of the boar sperm, our results cannot be extrapolated to other commercial farms, animals and conditions. Further studies involving more animals from different breeds, studs, management conditions and geographic locations will be needed for the accurate characterization of the boar sperm microbiome.

4.2. Sperm microbial composition

According to our data, the boar sperm microbiome differed from the profiles obtained on porcine gut where the most abundant phyla include *Bacteroidetes* and *Firmicutes* and the predominant genus are *Prevotella* and *Roseburia* [34]. On the contrary, our data highlights that in the porcine and in human sperm, the 4 most abundant phyla are coincident [10]. Moreover, 11 of the 20 most abundant genera in boar and human sperm were shared in both species. In human sperm, the most abundant bacteria were members of *Actinobacteria* (*Corynebacterium*), *Bacteroidetes* (*Prevotella*), *Firmicutes* (*Lactobacillus*, *Streptococcus*, *Staphylococcus*, *Planococcaceae*, *Finegoldia*), and *Proteobacteria* (*Haemophilus*, *Burkholderia*) [5]. The differences between the porcine and the human ejaculates could be attributed to multiple technical (e.g., the selection of antibiotics in extender and the removal of bacteria during the purification of the samples), environmental and biological causes. Although boar studs are kept in high hygienic conditions, pigs are in closer contact with surfaces, soil, faeces and water and are thus more exposed to environmental contaminants than humans.

405 The most abundant bacteria in the boar sperm are mostly environmental (*B.*
406 *megaterium* [35], *B. faecium* [36], *R. pickettii* [37]) and faecal (*C. hominis* [38]
407 and *E. coli*). This suggests that these bacteria have contaminated sperm after
408 ejaculation. *C. acnes* typically colonizes the human skin [39] but can be also
409 found in other compartments including the gastro-intestinal tract [40].
410 Interestingly, *B. subtilis*, a probiotic added in the pig feedstuff and allowed in
411 the European Union Register of Feed Additives, appeared as the 11th most
412 abundant bacteria in the boar sperm (Table 1), again suggesting that it
413 contaminated sperm after ejaculation.
414 Arkfen and co-authors [41] analyzed the airborne microbiome of hog farms and
415 found a similar composition of bacterial phyla as the one described in our study.
416 Moreover, our data is in line with the results obtained in other studies which
417 indicated that the bacteria present in sperm is a result of environmental
418 contamination, mostly attributed to prepuce fluid and hair [42], sinks and drains
419 in the farms, semen collection and processing utensils and the skin flora of
420 working staff [14].
421 Three ejaculates showed a much higher bacterial abundance when compared
422 to the average in all the samples (Fig. 1). Although we don't know the causes,
423 these elevated values of bacterial reads might have been caused by accidental
424 contamination of the ejaculate with particularly large chunks of environmental
425 debris present for example in the boar's prepuce or other surfaces.
426 **4.3. Pathogens and anti-microbial resistances**
427 We found several potential pathogens (Table 2) as included in the Professional
428 Pig Community pig333 site, The Pig Site and the Swine Health Information
429 Center. Some serotypes of these bacteria have been linked to diarrhea (*E. col*i,

C. difficile, *C. jejuni*), acute enteritis (*C. perfringens*) [43], botulism (*C. botulinum*), tuberculosis (*M. tuberculosis*) and enzootic pneumonia (*M. hyopneumoniae*) in swine [44]. While 4 of the 5 most abundant potential pathogens showed a continuous pattern of abundance across samples, *M. tuberculosis* was only present in 6 samples, all with moderate abundances (CPM between 28 and 84). This quasi bi-modal distribution cannot be explained by factors controlled in our study as these 6 pigs came from different farms, were of varying ages, their ejaculates were collected at different seasons of the year and there was thus no apparent link between these animals. The presence of *M. tuberculosis* complex has been already found in wild boar (*Sus scrofa*) in Eurasia [45]. The pathogenic potential of these bacterial species varies across strains depending on the presence of virulence factors and toxin production. Notwithstanding, our analysis does not allow concluding that any of the specimens identified in this study are pathogenic as the analysis did not have the power and specificity to detect the genes to discriminate between these serotypes.

In animal production systems, extended sperm is distributed to multiple farms and geographical locations and despite the fact that it is mixed with antibiotics, some bacteria remain in these ejaculates. Moreover, before they are inseminated into the sow, extended sperm doses will remain at 17 ºC in average up to few days, thus potentially allowing the selective growth of bacteria carrying ARGs. Therefore, ejaculates might be an important source and vehicle to disseminate these bacteria to other farms and animals. Hence, the vaginal microbiome in sows inseminated with these doses should be evaluated to determine how the sperm microbiome modulates the female tract,

how it impacts on the sow's health and fertility and the extent to which ARGs and pathogens are transmitted through artificial insemination.

We identified 3 ARGs that were present at CPM ≥ 1 in all the ejaculates (Table 2). These ARGs were point mutation variants in bacterial ribosomal RNA genes. The most abundant ARG potentially conferred resistance to *E. coli* to chloramphenicol, a broad-spectrum antibiotic predominantly active against gram negative bacteria used in human medicine but not authorized by the European Union for use in livestock. However, this antimicrobial can be synthesized by soil bacteria and it may thus be present in farms thereby allowing the generation of ARGs against it. Our results suggest a scarce presence of ARGs in our porcine sperm samples. The ejaculates were diluted with a commercial semen extender that contains the antibiotics apramycin, cephalosporin and gentamicin but no ARGs were found against these 3 antibiotics.

The 3 bacteria involved in these presumable ARGs (*E. coli*, *N. meningitidis* and *S. aureus*) were detected in our study but their abundances did not relate with the expression levels of their cognate ARGs (Fig. S1). The only exception is for *N. meningitidis* and the ARG for Spectinomycin (R^2 = 0.74), but this is largely due to one influential outlier ejaculate for which the abundance of both, these bacteria and ARG were remarkably elevated (Fig. S1). This indicates that not all the bacteria of these species carry the same load of ARG in each sample.

These results have to be taken as indicative as in this study we cannot conclude whether these abundances in CPM are large or modest. Moreover, the antimicrobial activity of these ARGs cannot be granted with our study. This activity should be confirmed with a classical microbiological analysis and

antimicrobial sensitivity testing with the target antibiotics, according to the Clinical & Laboratory Standards Institute (CLSI) guidelines [46, 47].

4.4. Relationship between the sperm microbiome and semen quality

As the microbiome is a complex system of microbial communities and its genomic characterization generates compositional and sparse data, we used an integrative approach that considers simultaneously the ejaculate bacterial composition and semen quality. This analysis led to the identification of two clusters with 30 and 10 samples each. Five traits and 13 bacterial features showed significant differences between the two clusters (Table S7 and Table 3). The fact that this analysis identified two categories based simultaneously on their semen quality and microbiome indicates that the two are related. VIAB_90 and ACRO_90 displayed stronger differences between the two groups than VIAB_5 and ACRO_5. This suggests that the long incubation favored the proliferation of bacteria and this led to a stronger bacterial impact on the phenotype. In farm conditions, most doses are used within 48 h after ejaculation but some may be kept up to 6 days. The impact of these conditions in the microbiome could be strong and it should be explored. Bacteria that remains active in the extended sperm despite the presence of antibiotics could alter sperm quality through several mechanisms including the competition for nutrients, the alteration of the microenvironment, the secretion of toxins, or the adhesion to the sperm cell membrane compromising sperm viability or aggregation. The 13 bacterial features showing differences between the two clusters included the genera *Acinetobacter*, *Stenotrophomonas* and *Rhodobacter* (Table 3). One study on human semen from Kiessling et al. [48] identified *Acinetobacter* bacteria in some of the semen samples that they

505 evaluated [48]. An *in vitro* study on rabbit sperm cultured under the presence
506 of *A. baumannii* showed that the motility of the spermatozoa was negatively
507 affected by the presence of this bacterium [49]. A study on boar sperm found
508 *A. iwoffi* in some samples and that the presence of this bacterium was
509 associated to higher production of Reactive Oxidative Species and lipid
510 peroxidation thus potentially altering some semen quality features [50].
511 *Stenotrophomonas* are also typically found in soil and plants and some
512 (including *S. maltophila*) can be opportunistic pathogens in humans. In swine,
513 it has been previously detected in sperm [13]. A case report on a dog with
514 conception failure and positive for *S. malthophila*, linked this bacteria with
515 semen quality [51]. Finally, the genus *Rhodobacter* includes several species
516 with a diverse range of energy-based metabolism but has not been previously
517 found in sperm nor linked to sperm quality. This genus can be found in varied
518 habitats including pig manure [52].
519 Semen quality is defined by a set of complex traits that depend on the genetics
520 and age of the boar and on multiple environmental factors including nutrition,
521 photoperiod and heat stress, housing conditions, semen collection frequency
522 and method, sperm dilution rate, storage media and packaging conditions [15].
523 In our study, we could not record most of these parameters. However, while
524 correcting the phenotypes by farm, age and season of the year, we indirectly
525 controlled for a proportion of these factors. First, all the ejaculates were
526 collected at night, stored under the same conditions and processed during the
527 following early afternoon. Moreover, housing conditions, nutrition, collection
528 method as well as storage conditions are farm specific and were thus indirectly
529 corrected when controlling by farm. The photoperiod and heat stress factors

were also indirectly considered as we also corrected the phenotypes by the season of the year. Nonetheless, we could not annotate the resting time (the time passed since the previous semen collection), a parameter that is known to affect semen quality [53]. Consequently, our results related to semen quality and the microbiome should be considered as indicative.

5. Conclusions

In conclusion, we have identified a large and varied population of bacteria contaminating the boar's extended sperm, including a small proportion of potential pathogens and ARGs. Moreover, some of these bacteria might be related to semen quality. This is of high relevance for two main reasons. First, these bacteria may affect sperm quality and male fertility. Second, since ejaculates are widely distributed across farms, they might be major disseminators of these microbes and ARGs. Thus, the microbial composition in the sperm of swine and other livestock species needs to be studied more profoundly. Moreover, we anticipate that in a not too distant future, the systematic microbiome analysis of semen ejaculates to identify the samples that contain potential pathogens will become common practice. At present, high throughput sequencing is still an expensive technology and this makes its routine application to assess semen quality in swine unfeasible. However, these costs are expected to keep decreasing in the years to come. This drop on sequencing costs should allow the systematic implementation of metagenomics to routinely assess the presence of pathogens and ARGs in the boar sperm.

555 **Acknowledgements**

556 We thank Sam Balasch (grup Gepork S.A.) and Craig Lewis (Genus - PIC) for
557 providing the blood and sperm samples.

558

559 **Funding**

560 This work was supported by the Spanish Ministry of Economy and
561 Competitiveness (MINECO) under grant AGL2013-44978-R and grant
562 AGL2017-86946-R and by the CERCA Programme/Generalitat de Catalunya.
563 AGL2017-86946-R was also funded by the Spanish State Research Agency
564 (AEI) and the European Regional Development Fund (ERDF). We thank the
565 Agency for Management of University and Research Grants (AGAUR) of the
566 Generalitat de Catalunya (Grant Numbers 2014 SGR 1528 and 2017 SGR
567 01060). We also acknowledge the support of the Spanish Ministry of Economy
568 and Competitivity for the Center of Excellence Severo Ochoa 2016–2019
569 (Grant Number SEV-2015-0533) grant awarded to the Centre for Research in
570 Agricultural Genomics (CRAG). MG acknowledges a Ph.D. studentship from
571 MINECO (Grant Number BES-2014-070560). YRC was funded by Marie
572 Skłodowska-Curie grant (P-Sphere) agreement No 6655919 (EU). LMZ is
573 recipient of a Ph.D. grant associated with the SEV-2015-0533 award to CRAG.

574

575 **Conflicts of interest**

576 The authors declare no conflict of interest.

577

578 **Availability of data**

The datasets generated and analysed are available at NCBI's BioProject PRJNA520978.

References

[1] Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. Nat Med. 2016;22:1079-89

[2] Mahboubi MA, Carmody LA, Foster BK, Kalikin LM, Vandevanter DR, Lipuma J. Cystic Fibrosis Respiratory Specimens. J Clin Microbiol. 2016;54:613-9

[3] Moreno I, Codoñer FM, Vilella F, Valbuena D, Martinez-Blanch JF, Jimenez-Almazán J, et al. Evidence that the endometrial microbiota has an effect on implantation success or failure. Am J Obstet Gynecol. 2016;215:684-703

[4] Weng SL, Chiu CM, Lin FM, Huang WC, Liang C, Yang T, et al. Bacterial communities in semen from men of infertile couples: Metagenomic sequencing reveals relationships of seminal microbiota to semen quality. PLoS ONE. 2014;9:e110152

[5] Baud D, Pattaroni C, Vulliemoz N, Castella V, Marsland BJ, Stojanov M. Sperm Microbiota and Its Impact on Semen Parameters. Front Microbiol. 2019;10:1-9

[6] Mändar R, Punab M, Borovkova N, Lapp E, Kiiker R, Korrovits P, et al. Complementary seminovaginal microbiome in couples. Res Microbiol. 2015;166:440-7

[7] Hou D, Zhou X, Zhong X, Settles M, Herring J, Wang L, et al. Microbiota of the seminal fluid from healthy and infertile men. Fertil Steril. 2013;100:1261-9

[8] Úbeda JL, Ausejo R, Dahmani Y, Falceto MV, Usan A, Malo C, et al. Adverse effects of members of the Enterobacteriaceae family on boar sperm quality. Theriogenology. 2013;80:565-70

[9] Sepúlveda L, Bussalleu E, Yeste M, Bonet S. Effect of Pseudomonas aeruginosa on sperm capacitation and protein phosphorylation of boar spermatozoa. Theriogenology. 2016;85:1421-31

[10] Swanson GM, Moskovtsev S, Librach C, Pilsner JR, Goodrich R, Krawetz SA. What human sperm RNA-Seq tells us about the microbiome. J Assist Reprod Gen. 2020;37:359-68

[11] Kuster CE, Althouse GC. The impact of bacteriospermia on boar sperm storage and reproductive performance. Theriogenology. 2016;85:21-6

[12] Even G, Mottais D, Morien F, Pham MD, Ostergaard A, Martel S, et al. Porcine bacteriospermia examined by high-throughput sequencing. Theriogenology. 2020;142:268-75

[13] Althouse GC, Lu KG. Bacteriospermia in extended porcine semen. Theriogenology. 2005;63:573-84

[14] Schulze M, Ammon C, Rüdiger K, Jung M, Grobbel M. Analysis of hygienic critical control points in boar semen production. Theriogenology. 2015;83:430-7

[15] Lopez Rodriguez A, Van Soom A, Arsenakis I, Maes D. Boar management and semen handling factors affect the quality of boar extended semen. Porcine Health Manag. 2017;3:15

[16] Schulze M, Schäfer J, Simmet C, Jung M, Gabler C. Detection and characterization of Lactobacillus spp. In the porcine seminal plasma and their influence on boar semen quality. PLoS ONE. 2018;13:1-16

[17] Gòdia M, Mayer FQ, Nafissi J, Castelló A, Rodríguez-Gil JE, Sánchez A, et al. A technical assessment of the porcine ejaculated spermatozoa for a sperm-specific RNA-seq analysis. Syst Biol Reprod Med. 2018;64:291-303
[18] Bamba K. Evaluation of Acrosomal Integrity of Boar Spermatozoa by Bright Field Microscopy Using an Eosin-Nigrosin Stain. Theriogenology. 1988;29:1245-51
[19] RodriguezGil JE, Rigau T. Effects of ouabain on the response to osmotic changes in dog and boar spermatozoa. Theriogenology. 1996;45:873-88
[20] Gòdia M, Estill M, Castelló A, Balasch S, Rodríguez-Gil JE, Krawetz SA, et al. A RNA-Seq Analysis to Describe the Boar Sperm Transcriptome and Its Seasonal Changes. Front Genet. 2019;10:299
[21] Gòdia M, Castelló A, Rocco M, Cabrera B, Rodríguez-Gil JE, Balasch S, et al. Identification of circular RNAs in porcine sperm and evaluation of their relation to sperm motility. Sci Rep. 2020;10:7985
[22] Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20
[23] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357-60
[24] Bao W, Kojima KK, Kohany O. Repbase Update, a database of eukaryotic repetitive elements. Mob DNA. 2015;6:11
[25] Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46
[26] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. BMC Bioinformatics. 2009;10:1-9
[27] Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017;45:D566-D73
[28] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2017.
[29] Yin P, Fan XT. Estimating R-2 shrinkage in multiple regression: A comparison of different analytical methods. J Exp Educ. 2001;69:203-24
[30] Zingaretti LM, Renand G, Morgavi DP, Ramayo-Caldas Y. Link-HD: a versatile framework to explore and integrate heterogeneous microbial communities. Bioinformatics. 2019;36:2298-9
[31] Paulson JN, Pop M, Bravo HC. metagenomeSeq: Statistical analysis for sparse high-throughput sequencing. Bioconductor package. 2013
[32] Althouse GC, Pierdon MS, Lu KG. Thermotemporal dynamics of contaminant bacteria and antimicrobials in extended porcine semen. Theriogenology. 2008;70:1317-23
[33] Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nat Rev Microbiol. 2018;16:410-22
[34] Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, et al. A reference gene catalogue of the pig gut microbiome. Nat Microbiol. 2016;1:16161
[35] Vary PS. Prime Time for Bacillus megaterium. Microbiology. 1994;140:1001-13
[36] Collins MD, Brown J, Jones D. Brachybacterium faecium gen. nov., sp. nov., a Coryneform Bacterium from Poultry Deep Litter. Int J Syst Evol Microbiol. 1988;38:45-8
[37] Labarca JA, Trick WE, Peterson CL, Carson LA, Holt SC, Arduino MJ, et al. A multistate nosocomial outbreak of Ralstonia pickettii colonization associated with an intrinsically contaminated respiratory care solution. Clin Infect Dis. 1999;29:1281-6

[38] Lawson AJ, Linton D, Stanley J. 16S rRNA gene sequences of 'Candidatus Campylobacter hominis', a novel uncultivated species, are found in the gastrointestinal tract of healthy humans. Microbiology. 1998;144:2063-71
[39] Brüggemann H, Henne A, Hoster F, Liesegang H, Wiezer A, Strittmatter A, et al. The Complete genome of Propionibacterium Acnes, a Commensal of Human Skin. Science. 2004;305:671-4
[40] Perry A, Lambert P. Propionibacterium acnes: infection beyond the skin. Expert Rev Anti Infect Ther. 2011;9:1149-56
[41] Arfken AM, Song B, Sung J-S. Comparison of airborne bacterial communities from a hog farm and spray field. J Microbiol Biotechnol. 2015;25:709-17
[42] Goldberg AM, Argenti LE, Faccin JE, Linck L, Santi M, Bernardi ML, et al. Risk factors for bacterial contamination during boar semen collection. Res Vet Sci. 2013;95:362-7
[43] Vidal A, Martin-Valls GE, Tello M, Mateu E, Martin M, Darwich L. Prevalence of enteric pathogens in diarrheic and non-diarrheic samples from pig farms with neonatal diarrhea in the North East of Spain. Vet Microbiol. 2019;237:108419
[44] Siqueira FM, Perez-Wohlfeil E, Carvalho FM, Trelles O, Schrank IS, Vasconcelos ATR, et al. Microbiome overview in swine lungs. PLoS ONE. 2017;12:e0181503
[45] Nol P, Ionescu R, Welearegay TG, Barasona JA, Vicente J, Beleno-Saenz KD, et al. Evaluation of Volatile Organic Compounds Obtained from Breath and Feces to Detect Mycobacterium tuberculosis Complex in Wild Boar (Sus scrofa) in Donana National Park, Spain. Pathogens. 2020;9:346
[46] CLSI. Performance Standards for Antimicrobial Susceptibility Testing. Wayne, PA, USA: Clinical and Laboratory Standards Institute; 2019.
[47] CLSI. Performance Standards for Antimicrobial Disk and Dilution Susceptibility Tests for Bacteria Isolated from Animals. Wayne, PA, USA: Clinical and Laboratory Standards Institute; 2018.
[48] Kiessling AA, Desmarais BM, Yin HZ, Loverde J, Eyre RC. Detection and identification of bacterial DNA in semen. Fertil Steril. 2008;90:1744-56
[49] Tvrdá E, Ďuračka M, Kántor A, Halenár M, Hleba L. *In vitro* effects of *Acinetobacter Baumannii* and selected natural biomolecules on rabbit spermatozoa motility. AGROFOR Int Journ. 2018;3:54-63
[50] Ďuračka M, Tvrda E. The presence of bacterial species in boar semen and their impact on the sperm quality and oxidative balance. J Anim Sci. 2018;96:501
[51] Domoslawska A, Zdunczyk S, Jurczak A, Janowski T. Stenotrophomonas maltophilia isolated from prostatic fluid as an infertility factor in a male dog. Andrologia. 2017;49:e12769
[52] Yang YY, Pereyra LP, Young RB, Reardon KF, Borch T. Testosterone-Mineralizing Culture Enriched from Swine Manure: Characterization of Degradation Pathways and Microbial Community Composition. Environ Sci Technol. 2011;45:6879-86
[53] Pruneda A, Pinart E, Dolors Briz M, Sancho S, Garcia-Gil N, Badia E, et al. Effects of a high semen-collection frequency on the quality of sperm from ejaculates and from six epididymal regions in boars. Theriogenology. 2005;63:2219-32

713 **Tables:**

714 **Table 1.** List of the 20 most abundant bacteria in the sperm from the 40 Pietrain boars.

715

Species	**Average abundance**	**Average percentage over all bacterial reads**	**Median abundance**	**CV**	**Maximum abundance**	**Minimum abundance**
Bacillus megaterium	867.51	4.31	827.93	0.85	4,026.18	0.00
Brachybacterium faecium	673.39	3.34	98.17	2.06	6,337.78	0.05
Bacillus coagulans	252.71	1.25	7.64	1.81	1,650.99	0.14
Campylobacter hominis	205.28	1.02	8.19	3.56	4,483.10	0.00
Psychrobacter sp PRwf-1	189.58	0.94	16.59	2.28	2,091.32	0.48
Cutibacterium acnes	154.74	0.77	87.77	1.81	1,771.66	8.76

Methylotenera mobilis	137.84	0.68	0.92	5.66	4,948.31	0.00
Escherichia coli	136.84	0.68	93.68	0.88	487.26	11.14
Porphyromonas asaccharolytica	136.19	0.68	4.72	3.48	2,614.58	0.00
Ralstonia pickettii	134.52	0.67	67.65	1.61	1,026.71	0.36
Bacillus subtilis	109.63	0.54	87.57	0.81	413.54	6.54
Acinetobacter baumannii	86.29	0.43	8.32	2.93	1,134.84	1.68
Thauera sp MZ1T	80.12	0.40	5.24	3.47	1,675.64	0.05
Saccharomonospora viridis	79.09	0.39	13.27	3.61	1,818.43	0.08
Anaerococcus prevotii	67.92	0.34	12.49	3.18	1,346.78	0.09
Aequorivita sublithincola	67.04	0.33	1.48	5.08	2,139.66	0.34
Advenella kashmirensis	61.93	0.31	3.48	2.87	994.21	0.00
Ornithobacterium rhinotracheale	60.96	0.30	1.44	5.55	2,145.52	0.02

Intrasporangium calvum	58.76	0.29	11.99	3.45	1,266.83	0.05
Pusillimonas sp T7-7	56.98	0.28	1.26	4.67	1,649.63	0.00

716

717 CV: Coefficient of variation.

718

719

720

721

722

723

724

725

726

727

728

729

730 **Table 2.** List of potential pathogens and antimicrobial resistance genes identified in the sperm from the 40 Pietrain boars.

Potential pathogen species	Average abundance	Median abundance	CV	Maximum abundance	Minimum abundance	Disease / health condition
Escherichia coli	136.84	93.68	0.88	487.26	11.14	Diarrhoea and high mortality in piglets
Clostridioides difficile	49.77	16.74	1.76	338.49	0.88	Diarrhoea in piglets
Clostridium perfringens	14.05	5.60	1.47	95.99	0.47	Chronic or acute enteritis in piglets. Sometimes also gangrene and sudden death in adults
Clostridium botulinum	7.67	2.92	1.71	69.73	0.33	Toxins produced by this bacteria cause a progressive flaccid paralysis, but pigs are very resistant to the toxin
Mycobacterium tuberculosis	7.46	0.00	2.60	84.45	0.00	Tuberculosis
Mycoplasma hyopneumoniae	6.24	5.48	0.78	23.39	0.82	Enzootic pneumonia
Campylobacter jejuni	5.86	0.05	3.11	90.93	0.00	Clinical signs are not always present but can cause a watery diarrhea with mucous and blood. Also, food-borne illness in humans
Staphylococcus aureus	4.71	3.23	1.16	32.5	0.2	Occasional cause of abscesses, arthritis, osteomyelitis, mastitis and skin conditions
Erysipelothrix rhusiopathiae	3.91	0.64	2.43	42.22	0.00	Erysipela: skin lesion and arthritis
[Haemophilus] parasuis	3.00	0.12	3.81	67.62	0.00	Glässer disease: polyserosistis and sporadic arthritis

Streptococcus suis	2.42	0.60	2.35	29.64	0.05	Streptococcal infection with pneumonia, septicemia, arthritis, etc. Zoonotic potential
Listeria monocytogenes	2.21	0.97	1.70	21.40	0.02	Rare systemic bacterial septicemia
Potential antibiotic resistant gene						
ARO:3003497_Neisseria_meningitidis_16S_rRNA_mutation_spectinomycin	27.85	21.95	1.11	189.86	2.31	
ARO:3004058_Staphylococcus_aureus_23S_rRNA_with_mutation_linezolid	125.07	103.89	0.78	400.27	5.16	
ARO:3004150_Escherichia_coli_23S_rRNA_with_mutation_chloramphenicol	316.72	198.07	1.01	1634.58	113.99	

CV: Coefficient of variation.

736 **Table 3.** List of bacteria displaying significant differences between clusters.

Bacterial feature	Global average	Average cluster 1	Average cluster 2	Fold change	P-value	Adjusted P-value
d__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rhodobacterales	130.18	161.53	36.13	1.13	5.21E-04	3.44E-02
d__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rhodobacterales|f__Rhodobacteraceae	129.64	160.94	35.72	1.13	5.02E-04	3.44E-02
d__Bacteria|p__Proteobacteria|c__Alphaproteobacteria|o__Rhodobacterales|f__Rhodobacteraceae|g__Rhodobacter	4.97	5.70	2.76	0.75	3.90E-04	3.44E-02
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pasteurellales	70.85	82.66	35.44	1.50	5.88E-04	3.44E-02
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pasteurellales|f__Pasteurellaceae	70.85	82.66	35.44	1.50	5.88E-04	3.44E-02
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales	977.88	973.87	989.92	1.65	6.08E-04	3.44E-02
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter	325.35	296.67	411.41	2.21	2.22E-05	5.45E-03
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_baumannii	86.29	65.97	147.25	2.21	4.68E-05	8.60E-03
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Pseudomonadales|f__Moraxellaceae|g__Acinetobacter|s__Acinetobacter_sp_ADP1	48.56	53.47	33.81	1.91	1.66E-04	2.20E-02

d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Xanthomonadales	42.32	47.94	25.49	1.24	6.61E-06	2.43E-03
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Xanthomonadales|f__Xanthomonadaceae	36.39	40.95	22.69	1.27	4.54E-06	2.43E-03
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Xanthomonadales|f__Xanthomonadaceae|g__Stenotrophomonas	6.85	6.93	6.61	1.11	2.10E-04	2.20E-02
d__Bacteria|p__Proteobacteria|c__Gammaproteobacteria|o__Xanthomonadales|f__Xanthomonadaceae|g__Stenotrophomonas|s__Stenotrophomonas_maltophilia	6.85	6.93	6.61	1.11	2.10E-04	2.20E-02

737 d: domain; p: phylum; c: class; o: order; f: family; g: genus; s: species.

Figure legends

Figure 1. Distribution of overall bacteria abundance for each animal.

Figure 2. Stackplot of the phyla distribution across the 40 sperm samples. The most abundant phyla were *Proteobacteria* followed by *Firmicutes*, *Actinobacteria* and *Bacteroidetes*.

Figure 3. Data structure from compromise configuration after applying a clustering using standard k-means with Link-HD. Cluster 1 (red) included 30 samples and cluster 2 (blue) 10 samples. Seven semen quality traits and 67 bacterial features were associated to this structure.

Supporting information

Supplementary Table S1. List of bacterial agents and diseases available at the Professional Pig Community pig333 site, the Pig Site and the Swine Health Information Center.

Supplementary Table S2. RNA-seq statistics for each of the 40 Pietrain samples. SD: Standard Deviation.

Supplementary Table S3. Full list of bacterial features and their abundances in the 40 Pietrain samples. CPM: Counts per Million reads; SD: Standard Deviation; d: domain; p: phylum; c: class; o: order; f: family; g: genus; s: species.

Supplementary Table S4. Phenotypic values for the 7 semen quality traits for each of the 40 samples. VIAB_5: percentage of viable sperm cells after 5 minutes of incubation at 37 ºC; VIAB_90: percentage of viable sperm cells after 90 min incubation

at 37 ºC; ACRO_5: percentage of cells with abnormal acrosomes after the 5 min; ACRO_90: 90 min incubation; ORT, percentage of viable cells after an osmotic stress (Osmotic Resistance Test); MT_5: percentage of motile cells after 5 min; MT_90: 90 min incubation.

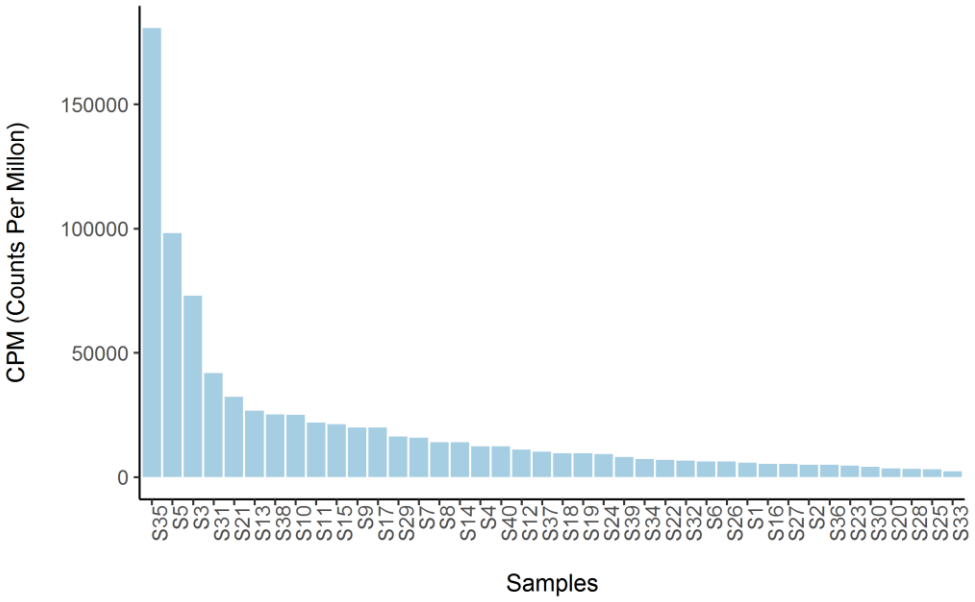
Supplementary Table S5. Detail of the samples ID belonging to each Link-HD cluster.

Supplementary Table S6. Detail of the traits and bacterial features contributing to the Link-HD compromise. VIAB_5: percentage of viable sperm cells after 5 minutes of incubation at 37 ºC; VIAB_90: percentage of viable sperm cells after 90 min incubation at 37 ºC; ACRO_5: percentage of cells with abnormal acrosomes after the 5 min; ACRO_90: 90 min incubation; ORT, percentage of viable cells after an osmotic stress (Osmotic Resistance Test); MT_5: percentage of motile cells after 5 min; MT_90: 90 min incubation; d: domain; p: phylum; c: class; o: order; f: family; g: genus; s: species.

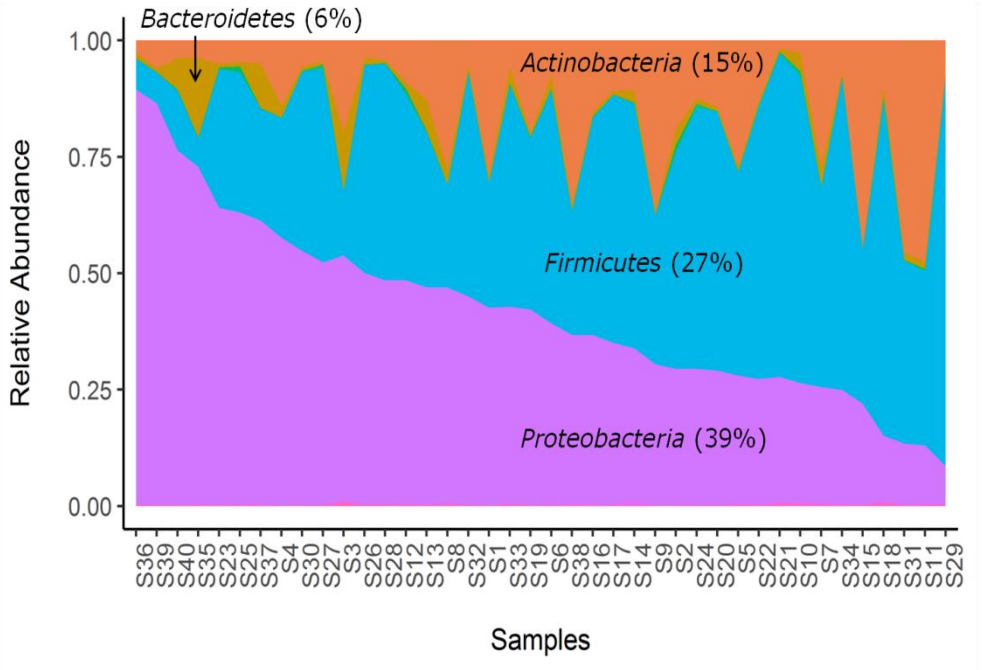
Supplementary Table S7. List of phenotypes displaying significant differences between the 2 clusters distinguishing both groups. SD: Standard deviation; VIAB_5: percentage of viable sperm cells after 5 minutes of incubation at 37 ºC; VIAB_90: percentage of viable sperm cells after 90 min incubation at 37 ºC; ACRO_5: percentage of cells with abnormal acrosomes after the 5 min; ACRO_90: 90 min incubation; ORT, percentage of viable cells after an osmotic stress (Osmotic Resistance Test); MT_5: percentage of motile cells after 5 min; MT_90: 90 min incubation.

Supplementary Figure S1. Linear regression plots (R^2) of the abundance of the antibiotic resistance genes (ARGs) and their related bacterial species.

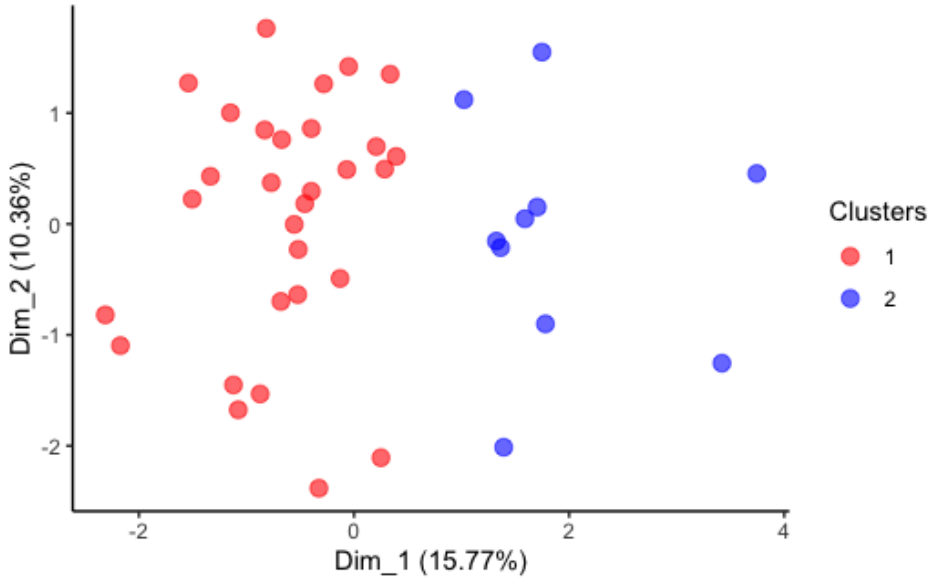
781



788 Figure 1.



789 Figure 2.



790

791 Figure 3.

792