1 **Title:**

2 A RNA-seq characterization of the porcine sperm microbiome

3

4 **Running Title:**

5 Boar sperm microbiome

6

7 M. Gòdia¹, Y. Ramayo-Caldas^{2,3}, L. M. Zingaretti¹, S. López⁴, J. E. Rodriguez-
8 Gil⁵, M. Yeste⁶, A. Sánchez^{1,7}, A. Clot^{1,8*}

9

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

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

15 3. GABI, INRA, AgroParisTech, Université Paris-Saclay, UMR 1313, Jouy-en-
16 Josas, France.

17 4. Department of condensed matter physics, Faculty of Physics 08028,
18 Universitat de Barcelona, Barcelona, Cataloia, Spain

19 5. Department of Animal Medicine and Surgery, School of Veterinary
20 Sciences, Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès,
21 Catalonia, Spain

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

26 7. Departament de Ciència Animal i dels Aliments, School of Veterinary
 27 Sciences, Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès,
 28 Catalonia, Spain

29 8. Consejo Superior de Investigaciones Científicas (CSIC), 08003, Barcelona,
 30 Catalonia, Spain

31

32 *Corresponding author:

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

34 **Summary**

35 The microbiome plays a key role in homeostasis and health and it has been
 36 also linked to fertility and semen quality in several animal species including
 37 swine. Despite the more than likely importance of sperm bacteria on the
 38 boar's reproductive ability and the dissemination of pathogens and
 39 antimicrobial resistance genes, a high throughput characterization of the
 40 swine sperm microbiome remains undone. The current study aimed at
 41 profiling the boar sperm bacterial population and its relationship with seven
 42 semen quality traits.

43 We carried RNA-seq on 40 ejaculates and we found that it contains a broad
 44 population of bacteria. The most abundant phyla were *Proteobacteria*
 45 (39.1%), *Firmicutes* (27.5%), *Actinobacteria* (14.9%) and *Bacteroidetes*
 46 (5.7%). The predominant species contaminated sperm after ejaculation from
 47 soil, faeces and water sources (*Bacillus megaterium*, *Brachybacterium*
 48 *faecium*, *Bacillus coagulans*). Some potential pathogens were also found but
 49 at relatively low levels (*Escherichia coli*, *Clostridioides difficile*, *Clostridium*
 50 *perfringens*, *Clostridium botulinum* and *Mycobacterium tuberculosis*). We also
 51 identified 3 potential antibiotic resistant genes from *E. coli* against
 52 chloramphenicol, *Neisseria meningitidis* against spectinomycin and
 53 *Staphylococcus aureus* against linezolid. None of these genes were highly
 54 abundant. Finally, we classified the ejaculates into categories according to
 55 their bacterial features and semen quality parameters and identified two
 56 categories that significantly differed for 5 semen quality traits and 13 bacterial
 57 features including the genera *Acinetobacter*, *Stenotrophomonas* and
 58 *Rhodobacter*. Our results show that boar semen contains a rich microbiome

59 with potential pathogens and antibiotic resistance genes which may affect its
60 reproductive performance.

61

62

63 **Keywords**

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

65 Introduction

66 Scientific research has lead to the discovery that microbiota, the
67 microorganisms present in the different compartments in an animal organism,
68 play a crucial role in physiological homeostasis and health (Schroeder and
69 Bäckhed, 2016, Mahboubi et al., 2016, Moreno et al., 2016) including sperm
70 quality and male fertility (Weng et al., 2014, Baud et al., 2019). The male's
71 reproductive ability is represented by a set of traits that are important for
72 human health and for the efficiency and sustainability of animal production. In
73 swine, semen quality is regularly measured in the artificial insemination studs
74 as a proxy of the fertilization ability of that sample. Growing research is being
75 devoted to understanding the biological basis and identifying molecular
76 markers linked to semen quality in humans and other animal species. As the
77 presence of bacterial communities in ejaculates is common and the
78 microbiome is popping up as a big contributor of a broad range of
79 phenotypes, several studies have been carried in the field of men fertility
80 (Mändar et al., 2015, Weng et al., 2014, Hou et al., 2013) and boar sperm
81 quality (Úbeda et al., 2013, Sepúlveda et al., 2016). Weng et al. (Weng et al.,
82 2014) identified a complex population of bacteria in human sperm but most
83 interestingly, found that the abundance of some bacteria was related to male
84 fertility. *Lactobacillus crispatus*, *Gardnerella vaginalis* and *Lactobacillus*
85 *acidophilus* were more abundant in the fertile samples whilst *Prevotella vibria*
86 and *Haemophilus parainfluenzae* were present at higher proportion in the
87 unfertile sperm (Weng et al., 2014). In a more recent study, a group led by
88 Stephen Krawetz (Swanson et al., 2020) used sperm RNA-seq datasets to
89 identify transcripts of bacterial origin and shed light to the bacterial

90 composition of an ejaculate. They found a diverse bacterial population mostly
 91 characterized by members of the phyla *Firmicutes*, *Proteobacteria*,
 92 *Bacteroidetes* and *Actinobacteria* (Swanson et al., 2020).

93 In pigs, the presence of bacteria in sperm is well documented and bacterial
 94 populations in ejaculates are common (Althouse and Lu, 2005). In pigs, most
 95 of the bacteria present in semen ejaculates seem to have an external origin
 96 and might thus have contaminated the sperm after ejaculation (Schulze et al.,
 97 2015). Ubeda and co-authors, using cell culture, concluded that the most
 98 abundant bacteria in pig semen were from the Enterobacteriaceae family and
 99 included, in order of abundance, *Serratia marcescens*, *Klebsiella oxytoca*,
 100 *Providencia stuartii*, *Morganella morganii*, *Proteus mirabilis*, and *Escherichia*
 101 *coli*. Among these, *S. marcescens*, *K. oxytoca*, *M. morganii*, or *P. mirabilis*
 102 were, in an individual basis, negatively associated with sperm quality (Úbeda
 103 et al., 2013). Schulze also recently identified the presence of several species
 104 of *Lactobacillus* and an association, *in vitro*, between the abundance of
 105 *Lactobacillus buchneri* with sperm motility, mitochondrial activity and
 106 membrane integrity and *Lactobacillus animalis* with motility (Schulze et al.,
 107 2018). To control bacterial growth in sperm, antimicrobial agents are
 108 commonly added as essential components to semen extenders (Althouse and
 109 Lu, 2005). Nonetheless, bacteria in these extended ejaculates can be still
 110 present due to incomplete efficiency of the antibiotics which could be partially
 111 caused by the expression of antimicrobial resistance genes (ARG) by these
 112 bacteria. Current high throughput sequencing technologies provide
 113 unprecedented capacity to study microbial ecosystems. However, it has not
 114 yet been applied to characterize the pig sperm microbiome. In the current

115 study, we characterized the boar sperm microbiome by RNA-seq on
116 ejaculates of 40 pigs and indicate the existence of a potential link between the
117 sperm microbiome and semen quality traits.

118

119 **Materials and methods**

120 **Sample collection**

121 Specialized professionals obtained fresh ejaculates from 40 Pietrain boars
122 from three different commercial farms, with the hand glove method. Ejaculates
123 were collected between March 2015 and January 2017 and boar ages ranged
124 from 9 to 55 months old. The fresh ejaculates were preserved in Androstar®
125 Plus boar semen extender that contains the following antibiotics: apramycin
126 sulphate, cephalosporin – third generation - and gentamicin sulphate. Sperm
127 ejaculates were phenotyped as in (Gòdia et al., 2018). Then, they were
128 purified using the BoviPure™ colloidal silica particles reagent (Nidacon;
129 Mölndal, Sweden) as detailed by Gòdia et al. to obtain only mature
130 spermatozoa cells (Gòdia et al., 2018).

131 **RNA extraction, qPCR validation, library prep, sequencing**

132 RNA from spermatozoa pellets was extracted using a standard Trizol®
133 approach and treated with TURBO DNA-free™ Kit (Invitrogen; Carlsbad,
134 USA) (Gòdia et al., 2018). RNA samples were subjected to RT-qPCR assays
135 to validate the absence of RNA from contaminating diploid cells (mainly
136 leukocytes and keratinocytes) and genomic DNA (Gòdia et al., 2018). Total
137 RNA was subjected to ribosomal RNA (rRNA) depletion with the Ribo-Zero
138 Gold rRNA Removal Kit (Illumina, CA, USA). RNA-seq libraries were prepared
139 with SMARTer Universal Low Input RNA library Prep kit (Clontech, France)

140 and sequenced in an Illumina's HiSeq2000/2500 system to generate 75 base
141 pair long paired end reads. Ten of these RNA-seq datasets have been
142 analysed to characterize the boar sperm transcriptome in our previous study
143 (Gòdia et al., 2019). The RNA-seq data used in this study (total RNA-seq
144 runs) is accessible at the NCBI's SRA within the under the SRA study
145 accession SRP183646.

146 **Identification of RNA molecules of bacterial origin**

147 RNA-seq reads of low quality and adaptor contaminations were removed with
148 Trimmomatic v.0.36 (Bolger et al., 2014). Filtered reads were then mapped to
149 the *Sus scrofa* genome (Sscrofa11.1) with HISAT2 v.2.1.0 (Kim et al., 2015)
150 with default parameters except "--max seeds 30" and "-k 2". The reads that
151 did not map to Sscrofa11.1 were screened against the catalogue of porcine
152 Transposable Elements from the Repbase database (Bao et al., 2015) with
153 HISAT2 v.2.1.0 (Kim et al., 2015). The reads that remained unmapped were
154 taxonomically classified and quantified with Kraken v.0.10.5 (Wood and
155 Salzberg, 2014) with a threshold score of 0.15 and using the default database
156 that includes NCBI taxonomic information and complete genomes from
157 RefSeq of archaeal, bacteria, phage and viral domains. Only the bacterial-
158 assigned reads were kept for further analysis. The number of reads assigned
159 to a given taxon was normalized by sequencing depth, as counts per million
160 (CPM). The list of potential pathogens in swine was extracted from the
161 Professional Pig Community pig333 site (www.pig333.com/pig-diseases). This
162 site includes 24 bacterial-borne diseases.

163 **Detection of antimicrobial resistance genes**

164 Unmapped reads were also subjected to identification and relative abundance
165 quantification of ARGs. ARGs were identified using BLASTN v.2.7.1
166 (Camacho et al., 2009) with 100% percentage identity using the
167 Comprehensive Antibiotic Resistance Database (CARD) v.3.0.0 (Jia et al.,
168 2017). The number of reads for each ARG was normalized by sequencing
169 depth, as CPM. The read coverage across ARGs of point mutations was
170 individually visualized using R v.3.5.3 (R Core Team, 2017).

171 **Relation between bacterial abundance and semen quality traits**

172 Seven sperm phenotypes were measured in the 40 samples as previously
173 described (Gòdia et al., 2018). These included the percentage of viable sperm
174 cells after 5 minutes of incubation at 37 °C (VIAB_5), the percentage of viable
175 sperm cells after 90 min incubation at 37 °C (VIAB_90), percentage of cells
176 with abnormal acrosomes after the 5 min (ACRO_5) and the 90 min
177 (ACRO_90) incubation, the percentage of motile cells measured with the
178 computer-assisted semen analysis (CASA) system (Integrated Sperm
179 Analysis System V1.0; Proiser) after 5 min (MT_5) and 90 min (MT_90)
180 incubation and the percentage of viable cells after an osmotic stress (ORT,
181 Osmotic Resistance Test). All these phenotypes were corrected using the R
182 function “lm” (R Core Team, 2017) for the fixed variables: farm, age and
183 season/year. To calculate the relationship between the semen quality traits
184 and the bacterial features of the sperm microbiome we used Link-HD
185 (Zingaretti et al., 2019), a tool for the integration of heterogeneous data that
186 considers the compositional nature of microbiome datasets. This analysis
187 included the 7 corrected phenotypes and the bacterial features with average

188 CPM ≥ 1 and representing more than 0.001% of the total bacterial read
189 counts.

190 The phenotypic values of the clusters obtained by Link-he abundances of the
191 733 bacterial features of the cluster compromises obtained by Link-HD were
192 then compared using the fitZig function of the metagenomeSeq package v.
193 1.28.2 (Paulson et al., 2013). This function uses a Zero-Inflated Gaussian
194 (Zig) distribution to model the log2-abundances of the Operational Taxonomic
195 Units OTUs) and it is calculated through an Expectation Maximisation (EM)
196 algorithm.

197

198 **Results**

199 **RNA-seq statistics**

200 We carried RNA-seq on 40 ejaculates each from a different Pietrain pig and
201 obtained an average of 40.7 M reads per sample. In average, 98.5% of the
202 reads passed the quality control and 82.7% mapped to the porcine genome
203 (Sscrofa11.1). A tiny proportion (0.012%) of the unmapped reads aligned to
204 Repbase (Bao et al., 2015) and 25.1% (an average of 1.7 M reads per library)
205 mapped to microbial genomes with Kraken (Table S1).

206 **Description of the boar sperm microbiome**

207 We identified 733 bacterial features with average abundance ≥ 1 CPM and
208 representing more than 0.001% of the total bacterial read counts. The total
209 bacterial abundance across samples varied between 2,241 and 180,624
210 CPMs (Fig. 1 and Table S2). The average and median abundances of
211 bacterial reads were 20,149 and 9,785 CPM, respectively and 3 ejaculates
212 had more than 70,000 bacterial CPM (Fig. 1). The bacterial features included

15 phyla (Fig. 2 and Table S2). 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 S2). At the species level, the analysis identified 254 bacterial species (Table S2). 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).

200 **Boar sperm safety: pathogens and antibiotic resistance genes**

We found 11 potentially pathogenic species of bacteria with average abundance ≥ 1 CPM and representing more than 0.001% of the total bacterial read counts, according to the Professional Pig Community pig333 site (www.pig333.com/pig-diseases) 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 bacteria species, with 137 and 50 CPM, respectively, these potential bugs were in general displaying low 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 S2).

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.*

238 *coli* 23S rRNA with mutation conferring resistance to chloramphenicol.

239 Moreover, all the samples presented CPM ≥ 1 for these 3 ARGs (Table 2).

240 **Relationship between the sperm microbiome and semen quality**

241 To identify potential relationships between bacterial abundances and semen
242 quality we employed Link-HD (Zingaretti et al., 2019), a recently developed
243 tool based on STATIS methodology to integrate heterogeneous datasets. This
244 approach analyzes different types of variables measured on the same
245 samples, here bacterial abundance and semen quality phenotypes. To the
246 end, the tool turns each raw data into cross-product matrix, computed on the
247 distances between samples, which are then combined in a common
248 configuration named compromise. A classical Principal Component Analysis
249 (PCA) decomposes the compromise variance into orthogonal components
250 and data structure can be easily recovered using standard clustering
251 techniques. In this study, the samples were clustered into categories
252 according to their microbiome and their semen quality. After quality controls,
253 we included the 733 bacterial features (from phyla to species in Table S2) and
254 7 semen quality traits (Table S3). Link-HD structured the purified ejaculates
255 into 2 clusters with 30 (cluster 1) and 10 (cluster 2) samples each (Fig. 3 and
256 Table S4). The analysis also recovers the contribution of each feature into the
257 common structure, which facilitates the interpretability of the results. We found
258 that the 7 semen traits and 67 of the 733 bacterial features associated with
259 the whole-compromise structure (Table S5).

260 We then compared the distribution of these 7 phenotypes and 67 bacterial
261 features in each cluster. The 2 categories showed statistically significant
262 differences for 5 traits. MT_5 and MT_90 did not differ between both groups

(Table 3). The feature abundances between the 2 clusters were compared. Thirteen bacterial features resulted in nominal significant differences between clusters (Table 4). These included the genus *Acinetobacter*, *Stenotrophomonas* and *Rhodobacter* (Table 4).

Discussion

Technical considerations

Bacterial composition and abundance in different animal compartments has been proven to affect a plethora of phenotypes in humans and domestic animals. This has become more obvious with the recent development of high throughput sequencing tools that allow the systematic characterization of microbiomes (Pennisi, 2007). In 2014, Weng et al. (Weng et al., 2014) identified a particular microbiome composition in human ejaculated sperm and related the abundance of several bacteria, most remarkably *Lactobacillus* and *Prevotella*, to semen quality. We carried RNA-seq on the extended sperm from 40 pigs after purification of the samples to remove somatic cells and bacteria with Bovipure™ with the aim to characterize the boar semen transcriptome in relation to sperm quality. After mapping the RNA-seq reads to the pig genome (Sscrofa11.1), between 9 and 31% of the reads remained unmapped (Table S1). We hypothesized that a proportion of the unmapped sequences could correspond to bacterial transcripts and aligned these reads to microbiome genomes. We identified a rich population of bacteria with a diverse abundance profile between the ejaculates. Despite the fact that we processed extended sperm containing antibiotics and that we treated these samples to remove micro-organisms, we found evidences of bacterial

288 presence in their sequenced RNAs. This indicates that the extender does not
 289 eliminate all the bacteria present in the ejaculate. It is even possible that these
 290 antibiotics promote a positive selection for the antibiotic resistant bacteria. In
 291 fact, we observed the presence of 3 ARGs that confer resistance to
 292 spectinomycin, linezolid and chloramphenicol. The results also suggest that a
 293 large amount of bacteria may be present in the sperm doses used in artificial
 294 insemination.

295 The bioinformatics tool Kraken (Wood and Salzberg, 2014) was designed to
 296 study meta-genomes and quantify the abundance of bacteria based in their
 297 DNA. Kraken has been also used to characterize the sperm microbiome using
 298 RNA-seq datasets in human (Swanson et al., 2020). While meta-genomics
 299 strictly focuses on the abundance of bacterial specimens, meta-
 300 transcriptomics informs on the expression of their genes and thus the function
 301 and activity of these micro-organisms in the sample. Our data provides a
 302 quantification of each bacterium based in the overall expression of their
 303 transcripts which accounts for both the bacterial abundance and their gene
 304 expression activity and have the additional advantage to account for active
 305 microorganisms. In other words, we cannot state without uncertainty whether
 306 one bacterium is more abundant than another in one sample but we can
 307 assume that this is the most likely scenario as in part, our measures are
 308 reflecting these abundances. For this reason and to ease the message
 309 provided in this manuscript, we have referred to bacterial abundance
 310 throughout the article.

311 **Sperm microbial composition**

312 According to our data, the boar sperm microbiome differed from the profiles
 313 obtained on porcine gut where the most abundant phyla include *Bacteroidetes*
 314 and *Firmicutes* and the predominant genus are *Prevotella* and *Roseburia*
 315 (Xiao et al., 2016). On the contrary, our data highlights that in the porcine and
 316 in human sperm, the 4 most abundant phyla are coincident (Swanson et al.,
 317 2020). Moreover, 11 of the 20 most abundant genera in boar and human
 318 sperm were shared in both species. In human sperm, the most abundant
 319 bacteria were members of *Actinobacteria* (*Corynebacterium*), *Bacteroidetes*
 320 (*Prevotella*), *Firmicutes* (*Lactobacillus*, *Streptococcus*, *Staphylococcus*,
 321 *Planococcaceae*, *Fingoldia*), and *Proteobacteria* (*Haemophilus*,
 322 *Burkholderia*) (Baud et al., 2019). The differences between the porcine and
 323 the human ejaculates could be attributed to multiple technical (e.g., the
 324 selection of antibiotics in extender and the removal of bacteria during the
 325 purification of the samples), environmental and biological causes. Although
 326 boar studs are kept in high hygienic conditions, pigs are in closer contact with
 327 surfaces, soil, faeces and water and are thus more exposed to environmental
 328 contaminants than humans.

329 The most abundant bacteria in the boar sperm are mostly environmental (*B.*
 330 *megaterium* (Vary, 1994), *B. faecium* (Collins et al., 1988), *R. pickettii*
 331 (Labarca et al., 1999)) and faecal (*C. hominis* (Lawson et al., 1998) and *E.*
 332 *coli*). This suggests that these bacteria have contaminated sperm after
 333 ejaculation. *C. acnes* typically colonizes the human skin (Brüggemann et al.,
 334 2004) but can be also found in other compartments including the gastro-
 335 intestinal tract (Perry and Lambert, 2011). Interestingly, *B. subtilis*, a probiotic
 336 added in the pig feedstuff and allowed in the European Union Register of

337 Feed Additives, appeared as the 11th most abundant bacteria in the boar
338 sperm (Table 1), again suggesting that it contaminated sperm after
339 ejaculation.

340 Arkfen and co-authors (Arkfen et al., 2015) analysed the airborne microbiome
341 of hog farms and found a similar composition of bacterial phyla as the one
342 described in our study. Moreover, our data is in line with the latest
343 comprehensive study carried by Schulze et al. (Schulze et al., 2015) which
344 indicated that the major sources of contamination were of non-animal origin
345 and were mostly attributed to sinks and drains but also to other utensils and
346 the skin flora of working staff. This further supports the hypothesis that most
347 of the bacteria present in the pig sperm are from environmental
348 contamination.

349 Three ejaculates showed a much higher bacterial abundance when compared
350 to the average in all the samples (Fig. 1). Although we don't know the causes,
351 these elevated values of bacterial reads might have been caused by
352 accidental contamination of the ejaculate with particularly large chunks of
353 environmental debris present for example in the boar's prepuce or other
354 surfaces.

355 **Pathogens and anti-microbial resistances**

356 We found several potential pathogens (Table 2) as included in the
357 Professional Pig Community pig333 site (www.pig333.com/pig-diseases).
358 Some serotypes of these bacteria have been linked to diarrhea (*E. coli*, *C.*
359 *difficile*, *C. jejuni*), acute enteritis (*C. perfringens*), botulism (*C. botulinum*),
360 tuberculosis (*M. tuberculosis*) and enzootic pneumonia (*M. hyopneumoniae*)
361 in swine. 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, were collected at different seasons of the year and there was thus no apparent link between these animals. 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 °C 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.

The evaluation of the vaginal microbiome in sows inseminated with these doses should be evaluated to determine how the sperm microbiome modulates the vaginal counterpart and 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 mutations in bacterial ribosomal RNA

387 genes. The most abundant ARG potentially conferred resistance to *E. coli* to
 388 chloramphenicol, a broad spectrum antibiotic predominantly active against
 389 gram negative bacteria used in human medicine but not authorized by the
 390 European Union for use in livestock. However, this antimicrobial can be
 391 synthesized by soil bacteria and it may thus be present in farms thereby
 392 allowing the generation of ARGs against it. Our results suggest a scarce
 393 presence of ARGs in our porcine sperm samples. The ejaculates were diluted
 394 with a commercial semen extender that contains the antibiotics apramycin,
 395 cephalosporin and gentamicin but no ARGs were found against these 3
 396 antibiotics.

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

405 These results have to be taken as indicative as in this study we cannot
 406 conclude whether these abundances in CPM are large or modest. Moreover,
 407 the antimicrobial activity of these ARGs cannot be granted and should be
 408 confirmed with a microbiological analysis by growing the bacteria in media
 409 containing the target antibiotic using, for example, the agar disc diffusion
 410 method. The appearance and spread of antibiotic resistant bacteria is one of
 411 the biggest present and future challenges of our society due to their impact on

412 animal, environmental and human health. These resistances are growing and
 413 jeopardizing the treatments to fight bacterial infections. Recently, a resistance
 414 against collistin, a latest resource antibiotic used to eliminate *C. difficile* from
 415 hospitals appeared in a pig farm in China (Collins and Riley, 2018). This ARG
 416 disseminated quickly and reached human hospitals and livestock farms
 417 throughout the world. *C. difficile* is a bacteria difficult to treat that threatens
 418 some patients and have fatal consequences (Collins and Riley, 2018). In the
 419 framework of One Health, actors from multiple sectors (animal breeders,
 420 researchers and policy makers) are seriously studying to reduce the use of
 421 antibiotics to decrease the rate of development of ARGs. In this sense, as
 422 happens with the dissemination of pathogen bacteria, ejaculates can be
 423 responsible for a significant proportion of the dissemination of ARGs between
 424 farms.

425 **Relationship between the sperm microbiome and semen quality**

426 Motivated by the results from Weng et al. that found a relationship between
 427 male fertility and the microbiome of the human sperm (Weng et al., 2014) we
 428 sought to assess whether there was a relation between the sperm microbiome
 429 and semen quality. Studies in pig have found detrimental associations
 430 between *E. coli* and sperm cell agglutination and reduced fertility (Maroto
 431 Martin et al., 2010), *C. perfringens* and diminished sperm motility and viability
 432 (Sepúlveda et al., 2013) and *Pseudomonas aureginosa* and decreased
 433 longevity and fertilizing ability (Sepúlveda et al., 2014).

434 The microbiome is a complex system of microbial communities and its
 435 genomic characterization generates compositional and sparse data. Thus, we
 436 decided to treat the sperm bacterial communities holistically when assessing

its relation to semen quality instead of analyzing the individual impact of each bacterial entity. For this, 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 3 and Table 4). Semen quality was worst for cluster 2 (Table 3). Larger values for VIAB and ORT indicate more robust and viable semen whilst higher values for ACRO indicate a larger proportion of spermatozoa with deficient acrosomes.

The fact that this analysis identified two clear categories based simultaneously on their semen quality and microbiome indicates that the two are related. Interestingly, 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 lead to a stronger bacterial impact on the phenotype. In farm conditions, extended ejaculates are kept at 17 °C until they are used for insemination. 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. The 13 bacterial features showing differences between the two clusters included the genera *Acinetobacter*, *Stenotrophomonas* and *Rhodobacter* (Table 4). *Acinetobacter* are typically found in soil, although *A. baumannii* can be also detected in hospitals where it can act as an opportunistic pathogen affecting immunocompromised patients. One study on human semen from Kiessling et al. (Kiessling et al., 2008) identified *Acinetobacter* bacteria in some of the semen samples that they evaluated

(Kiessling et al., 2008). An *in vitro* study on rabbit sperm cultured under the presence of *A. baumannii* showed that the motility of the spermatozoa was negatively affected by the presence of this bacterium (Tvrdá et al., 2018). A study on boar sperm found *A. iwoffi* in some samples and that the presence of this bacteria was associated to higher production of Reactive Oxidative Species and lipid peroxidation thus potentially altering some semen quality features not measured in that study (Ďuračka and Tvrdá, 2018). *Stenotrophomonas* are also typically found in soil and plants and some (including *S. maltophilia*) can be opportunistic pathogens in humans. In swine, it has been previously detected in sperm (Althouse and Lu, 2005). A case report on a dog with conception failure and positive for *S. maltophilia*, linked this bacteria with semen quality (Domoslawska et al., 2017). Finally, the genus *Rhodobacter* includes several species with a diverse range of energy-based metabolism but has not been previously found in sperm nor linked to sperm quality. This genus can be found in varied habitats including pig manure (Yang et al., 2011).

478

479 **Conclusions**

480 In conclusion, we have identified a large and varied population of bacteria
481 contaminating the boar's extended sperm, including a small proportion of
482 potential pathogens and ARGs. Moreover, some of these bacteria, such as
483 *Acinetobacter*, *Stenotrophomonas* and *Rhodobacter*, seem to be related to
484 semen quality. This is of high relevance for two main reasons. First, these
485 bacteria may affect sperm quality and male fertility. Second, since ejaculates
486 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 in sequencing costs should allow the systematic implementation of metagenomics to routinely assess the presence of pathogens and ARGs in the boar sperm.

Acknowledgements

We thank Sam Balasch (grup Gepork S.A.) and Craig Lewis (Genus - PIC) for providing the blood and sperm samples. This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) under grant AGL2013-44978-R and grant AGL2017-86946-R and by the CERCA Programme/Generalitat de Catalunya. AGL2017-86946-R was also funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERDF). We thank the Agency for Management of University and Research Grants (AGAUR) of the Generalitat de Catalunya (Grant Numbers 2014 SGR 1528 and 2017 SGR 01060). We also acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the Center of Excellence Severo Ochoa 2016–2019 (Grant Number SEV-2015-0533) grant awarded to the Centre for Research in

512 Agricultural Genomics (CRAG). MG acknowledges a Ph.D. studentship from
513 MINECO (Grant Number BES-2014-070560). YRC was funded by Marie
514 Skłodowska-Curie grant (P-Sphere) agreement No 6655919 (EU). LMZ is
515 recipient of a Ph.D. grant associated with the SEV-2015-0533 award to
516 CRAG.

517

518 Availability of data

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

521

522 References

- 523 Althouse, G. C. & Lu, K. G. (2005). Bacteriospermia in extended porcine
524 semen. *Theriogenology* **63**, 573-84.
- 525 Arfken, A. M., Song, B. & Sung, J.-S. (2015). Comparison of airborne bacterial
526 communities from a hog farm and spray field. *Journal of Microbiology*
527 *and Biotechnology* **25**, 709-17.
- 528 Bao, W., Kojima, K. K. & Kohany, O. (2015). Repbase Update, a database of
529 eukaryotic repetitive elements. *Mobile DNA* **6**, 11.
- 530 Baud, D., Pattaroni, C., Vulliemoz, N., Castella, V., Marsland, B. J. &
531 Stojanov, M. (2019). Sperm Microbiota and Its Impact on Semen
532 Parameters. *Frontiers in Microbiology* **10**, 1-9.
- 533 Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: A flexible trimmer
534 for Illumina sequence data. *Bioinformatics* **30**, 2114-20.
- 535 Brüggemann, H., Henne, A., Hoster, F., Liesegang, H., Wiezer, A.,
536 Strittmatter, A., Hujer, S. & Du, P. (2004). The Complete genome of
537 *Propionibacterium Acnes*, a Commensal of Human Skin. *Science* **305**,
538 671-4.
- 539 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer,
540 K. & Madden, T. L. (2009). BLAST+: Architecture and applications.
541 *BMC Bioinformatics* **10**, 1-9.

- 542 Collins, D. A. & Riley, T. V. (2018). Clostridium difficile in Asia: Opportunities
543 for One Health Management. *Tropical Medicine and Infectious Disease*
544 **4**, 7.
- 545 Collins, M. D., Brown, J. & Jones, D. (1988). Brachybacterium faecium gen.
546 nov., sp. nov., a Coryneform Bacterium from Poultry Deep Litter.
547 *International Journal of Systematic Bacteriology* **38**, 45-8.
- 548 Domoslawska, A., Zdunczyk, S., Jurczak, A. & Janowski, T. (2017).
549 Stenotrophomonas maltophilia isolated from prostatic fluid as an
550 infertility factor in a male dog. *Andrologia* **49**, e12769.
- 551 Ďuračka, M. & Tvrdá, E. (2018). The presence of bacterial species in boar
552 semen and their impact on the sperm quality and oxidative balance.
553 *Journal of Animal Science* **96**, 501.
- 554 Gòdia, M., Estill, M., Castelló, A., Balasch, S., Rodríguez-Gil, J. E., Krawetz,
555 S. A., Sánchez, A. & Clot, A. (2019). A RNA-Seq Analysis to Describe
556 the Boar Sperm Transcriptome and Its Seasonal Changes. *Front Genet*
557 **10**, 299.
- 558 Gòdia, M., Mayer, F. Q., Nafissi, J., Castelló, A., Rodríguez-Gil, J. E.,
559 Sánchez, A. & Clot, A. (2018). A technical assessment of the porcine
560 ejaculated spermatozoa for a sperm-specific RNA-seq analysis.
561 *Systems Biology in Reproductive Medicine* **64**, 291-303.
- 562 Hou, D., Zhou, X., Zhong, X., Settles, M., Herring, J., Wang, L., Abdo, Z.,
563 Forney, L. J. & Xu, C. (2013). Microbiota of the seminal fluid from
564 healthy and infertile men. *Fertility and Sterility* **100**, 1261-9.
- 565 Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K.,
566 Lago, B. A., Dave, B. M., Pereira, S., Sharma, A. N., Doshi, S., Courtot,
567 M., Lo, R., Williams, L. E., Frye, J. G., Elsayegh, T., Sardar, D.,
568 Westman, E. L., Pawlowski, A. C., Johnson, T. A., Brinkman, F. S. L.,
569 Wright, G. D. & McArthur, A. G. (2017). CARD 2017: Expansion and
570 model-centric curation of the comprehensive antibiotic resistance
571 database. *Nucleic Acids Research* **45**, D566-D73.
- 572 Kiessling, A. A., Desmarais, B. M., Yin, H. Z., Loverde, J. & Eyre, R. C.
573 (2008). Detection and identification of bacterial DNA in semen. *Fertility*
574 *and Sterility* **90**, 1744-56.

575 Kim, D., Langmead, B. & Salzberg, S. L. (2015). HISAT: a fast spliced aligner
576 with low memory requirements. *Nature Methods* **12**, 357-60.

577 Labarca, J. A., Trick, W. E., Peterson, C. L., Carson, L. A., Holt, S. C.,
578 Arduino, M. J., Meylan, M., Mascola, L. & Jarvis, W. R. (1999). A
579 multistate nosocomial outbreak of *Ralstonia pickettii* colonization
580 associated with an intrinsically contaminated respiratory care solution.
581 *Clinical infectious diseases : an official publication of the Infectious*
582 *Diseases Society of America* **29**, 1281-6.

583 Lawson, A. J., Linton, D. & Stanley, J. (1998). 16S rRNA gene sequences of
584 'Candidatus *Campylobacter hominis*', a novel uncultivated species, are
585 found in the gastrointestinal tract of healthy humans. *Microbiology* **144**,
586 2063-71.

587 Mahboubi, M. A., Carmody, L. A., Foster, B. K., Kalikin, L. M., Vandevanter,
588 D. R. & Lipuma, J. (2016). Cystic Fibrosis Respiratory Specimens.
589 *Journal of clinical microbiology* **54**, 613-9.

590 Mändar, R., Punab, M., Borovkova, N., Lapp, E., Kiiker, R., Korrovits, P.,
591 Metspalu, A., Krjutškov, K., Nlvak, H., Preem, J. K., Oopkaup, K.,
592 Salumets, A. & Truu, J. (2015). Complementary seminovaginal
593 microbiome in couples. *Research in Microbiology* **166**, 440-7.

594 Maroto Martin, L. O., Munoz, E. C., De Cupere, F., Van Driessche, E.,
595 Echemendia-Blanco, D., Rodriguez, J. M. & Beeckmans, S. (2010).
596 Bacterial contamination of boar semen affects the litter size. *Anim*
597 *Reprod Sci* **120**, 95-104.

598 Moreno, I., Codoñer, F. M., Vilella, F., Valbuena, D., Martinez-Blanch, J. F.,
599 Jimenez-Almazán, J., Alonso, R., Alamá, P., Remohí, J., Pellicer, A.,
600 Ramon, D. & Simon, C. (2016). Evidence that the endometrial
601 microbiota has an effect on implantation success or failure. *American*
602 *Journal of Obstetrics and Gynecology* **215**, 684-703.

603 Paulson, J. N., Pop, M. & Bravo, H. C. (2013). metagenomeSeq: Statistical
604 analysis for sparse high-throughput sequencing. *Bioconductor*
605 *package*.

606 Pennisi, E. (2007). METAGENOMICS: Massive Microbial Sequence Project
607 Proposed. *Science* **315**, 1781a.

- 608 Perry, A. & Lambert, P. (2011). Propionibacterium acnes: infection beyond the
609 skin. *Expert Review of Anti-infective Therapy* **9**, 1149-56.
- 610 R Core Team. 2017. *R: A language and environment for statistical computing*.
611 R Foundation for Statistical Computing, Vienna, Austria.
- 612 Schroeder, B. O. & Bäckhed, F. (2016). Signals from the gut microbiota to
613 distant organs in physiology and disease. *Nature Medicine* **22**, 1079-
614 89.
- 615 Schulze, M., Ammon, C., Rüdiger, K., Jung, M. & Grobbel, M. (2015).
616 Analysis of hygienic critical control points in boar semen production.
617 *Theriogenology* **83**, 430-7.
- 618 Schulze, M., Schäfer, J., Simmet, C., Jung, M. & Gabler, C. (2018). Detection
619 and characterization of Lactobacillus spp. In the porcine seminal
620 plasma and their influence on boar semen quality. *PLoS ONE* **13**, 1-16.
- 621 Sepúlveda, L., Bussalleu, E., Yeste, M. & Bonet, S. (2014). Effects of different
622 concentrations of Pseudomonas aeruginosa on boar sperm quality.
623 *Animal Reproduction Science* **150**, 96-106.
- 624 Sepúlveda, L., Bussalleu, E., Yeste, M. & Bonet, S. (2016). Effect of
625 Pseudomonas aeruginosa on sperm capacitation and protein
626 phosphorylation of boar spermatozoa. *Theriogenology* **85**, 1421-31.
- 627 Sepúlveda, L., Bussalleu, E., Yeste, M., Torner, E. & Bonet, S. (2013). How
628 do different concentrations of Clostridium perfringens affect the quality
629 of extended boar spermatozoa? *Animal Reproduction Science* **140**, 83-
630 91.
- 631 Swanson, G. M., Moskovtsev, S., Librach, C., Pilsner, J. R., Goodrich, R. &
632 Krawetz, S. A. (2020). What human sperm RNA-Seq tells us about the
633 microbiome. *Journal of Assisted Reproduction and Genetics* **37**, 359-
634 68.
- 635 Tvrdá, E., Ďuračka, M., Kántor, A., Halenár, M. & Hleba, L. (2018). *In vitro*
636 effects of Acinetobacter Baumannii and selected natural biomolecules
637 on rabbit spermatozoa motility. *AGROFOR International Journal* **3**, 54-
638 63.
- 639 Úbeda, J. L., Ausejo, R., Dahmani, Y., Falceto, M. V., Usan, A., Malo, C. &
640 Perez-Martinez, F. C. (2013). Adverse effects of members of the

641 Enterobacteriaceae family on boar sperm quality. *Theriogenology* **80**,
642 565-70.

643 Vary, P. S. (1994). Prime Time for *Bacillus megaterium*. *Microbiology* **140**,
644 1001-13.

645 Weng, S. L., Chiu, C. M., Lin, F. M., Huang, W. C., Liang, C., Yang, T., Yang,
646 T. L., Liu, C. Y., Wu, W. Y., Chang, Y. A., Chang, T. H. & Huang, H. D.
647 (2014). Bacterial communities in semen from men of infertile couples:
648 Metagenomic sequencing reveals relationships of seminal microbiota
649 to semen quality. *PLoS ONE* **9**, e110152.

650 Wood, D. E. & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic
651 sequence classification using exact alignments. *Genome Biology* **15**,
652 R46.

653 Xiao, L., Estellé, J., Kiellerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., Liang,
654 S., Pedersen, A. O., Kjeldsen, N. J., Liu, C., Maguin, E., Dore, J., Pons,
655 N., Le Chatelier, E., Prifti, E., Li, J., Jia, H., Liu, X., Xu, X., Ehrlich, S.
656 D., Madsen, L., Kristiansen, K., Rogel-Gaillard, C. & Wang, J. (2016). A
657 reference gene catalogue of the pig gut microbiome. *Nat Microbiol* **1**,
658 16161.

659 Yang, Y. Y., Pereyra, L. P., Young, R. B., Reardon, K. F. & Borch, T. (2011).
660 Testosterone-Mineralizing Culture Enriched from Swine Manure:
661 Characterization of Degradation Pathways and Microbial Community
662 Composition. *Environmental Science & Technology* **45**, 6879-86.

663 Zingaretti, L. M., Renand, G., Morgavi, D. P. & Ramayo-Caldas, Y. (2019).
664 Link-HD: a versatile framework to explore and integrate heterogeneous
665 microbial communities. *Bioinformatics*, 1-2.

Tables:

Table 1. List of the 20 most abundant bacteria in the boar sperm.

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

<i>Cutibacterium acnes</i>	154.74	0.77	87.77	1.81	1,771.66	8.76
<i>Methylobacterium mobilis</i>	137.84	0.68	0.92	5.66	4,948.31	0.00
<i>Escherichia coli</i>	136.84	0.68	93.68	0.88	487.26	11.14
<i>Porphyromonas asaccharolytica</i>	136.19	0.68	4.72	3.48	2,614.58	0.00
<i>Ralstonia pickettii</i>	134.52	0.67	67.65	1.61	1,026.71	0.36
<i>Bacillus subtilis</i>	109.63	0.54	87.57	0.81	413.54	6.54
<i>Acinetobacter baumannii</i>	86.29	0.43	8.32	2.93	1,134.84	1.68
<i>Thauera</i> sp MZ1T	80.12	0.40	5.24	3.47	1,675.64	0.05
<i>Saccharomonospora viridis</i>	79.09	0.39	13.27	3.61	1,818.43	0.08
<i>Anaerococcus prevotii</i>	67.92	0.34	12.49	3.18	1,346.78	0.09
<i>Aequorivita sublithicola</i>	67.04	0.33	1.48	5.08	2,139.66	0.34
<i>Advenella kashmirensis</i>	61.93	0.31	3.48	2.87	994.21	0.00

<i>Ornithobacterium rhinotracheale</i>	60.96	0.30	1.44	5.55	2,145.52	0.02
<i>Intrasporangium calvum</i>	58.76	0.29	11.99	3.45	1,266.83	0.05
<i>Pusillimonas sp T7-7</i>	56.98	0.28	1.26	4.67	1,649.63	0.00

CV: Coefficient of variation.

Table 2. List of potential pathogens and antimicrobial resistance genes identified in pig sperm.

Potential pathogen species	Average abundance	Median abundance	CV	Maximum abundance	Minimum abundance	Disease / health condition
<i>Escherichia coli</i>	136.84	93.68	0.88	487.26	11.14	Diarrhoea and high mortality in piglets
<i>Clostridioides difficile</i>	49.77	16.74	1.76	338.49	0.88	Diarrhoea in piglets
<i>Clostridium perfringens</i>	14.05	5.60	1.47	95.99	0.47	Chronic or acute enteritis in piglets. Sometimes also gangrene and sudden death in adults
<i>Clostridium botulinum</i>	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
<i>Mycobacterium tuberculosis</i>	7.46	0.00	2.60	84.45	0.00	Tuberculosis
<i>Mycoplasma hyopneumoniae</i>	6.24	5.48	0.78	23.39	0.82	Enzootic pneumonia
<i>Campylobacter jejuni</i>	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
<i>Erysipelothrix rhusiopathiae</i>	3.91	0.64	2.43	42.22	0.00	Erysipela: skin lesion and arthritis
<i>[Haemophilus] parasuis</i>	3.00	0.12	3.81	67.62	0.00	Glässer disease: polyserosistis and sporadic arthritis

<i>Streptococcus suis</i>	2.42	0.60	2.35	29.64	0.05	Streptococcal infection with pneumonia, septicemia, arthritis, etc. Zoonotic potential
<i>Listeria monocytogenes</i>	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.

Table 3. List of phenotypes displaying significant differences between the 2 clusters distinguishing both groups.

Parameter	VIAB_5	VIAB_90	ACRO_5	ACRO_90	ORT	MT_5	MT_90
Average cluster 1	0.39	3.78	-0.71	-2.00	3.81	-1.28	-0.58
Average cluster 2	-3.92	-14.59	5.16	9.31	-10.24	0.94	-3.06
Average all	-0.69	-0.81	0.76	0.83	0.30	-0.72	-1.20
SD cluster 1	3.50	6.38	3.37	3.62	4.55	17.48	21.12
SD cluster 2	3.78	11.02	5.28	9.53	8.61	13.81	26.37
Total SD	4.00	11.10	4.63	7.44	8.40	16.50	22.21
Mann Whitney U test (P-value)	0.003	1.1E-05	0.004	1.5E-04	2.3E-05	0.863	1.000

SD: Standard deviation; VIAB_5: percentage of viable sperm cells after 5 minutes of incubation at 37 °C; VIAB_90: percentage of viable sperm cells after 90 min incubation at 37 °C; 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.

Table 4. 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_bauman	86.29	65.97	147.25	2.21	4.68E-05	8.60E-03

nii						
d__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Pseudomonadales f__Moraxellaceae g__Acinetobacter s__Acinetobacter_sp_AD1	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

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

Figures:

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

Figure 2. Stackplot of the phyla distribution across the 40 sperm samples.

Figure 3. Data structure from compromise configuration after applying a clustering using standard k-means with Link-HD.

Supporting information

Supplementary Table S1. RNA-seq statistics for each of the 40 samples.

SD: Standard Deviation.

Supplementary Table S2. Full list of bacteria taxa and its abundance in the 40 samples.

CPM: Counts per Million reads.

Supplementary Table S3. Phenotypic values for the 7 semen quality traits for each of the 40 samples.

Supplementary Table S4. Detail of the samples belonging to each Link_HD cluster.

Supplementary Table S5. Detail of the traits and bacterial features contributing to the Link-HD compromise.

Supplementary Figure S1. Linear regression plots of the abundance of the ARGs and their related bacterial species.