Porcine bacteriospermia examined by high-throughput sequencing

Gaël Even, David Mottais, Florianne Morien, Minh Duc Pham, Anaïs Ostergaard, Sophie Martel, Sophie Merlin, Christophe Audebert

► To cite this version:

Gaël Even, David Mottais, Florianne Morien, Minh Duc Pham, Anaïs Ostergaard, et al.. Porcine bacteriospermia examined by high-throughput sequencing. *Theriogenology*, 2020, 142, pp.268 - 275. 10.1016/j.theriogenology.2019.10.034 . hal-03488561

HAL Id: hal-03488561

<https://hal.science/hal-03488561>

Submitted on 21 Dec 2021

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

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

Distributed under a Creative Commons Attribution - NonCommercial| 4.0 International License

Porcine bacteriospermia examined by high-throughput sequencing

Gaël Even^{a,b}, David Mottais^a, Florianne Morien^a, Minh Duc Pham^a, Anaïs Ostergaard^a, Sophie Martel^{a,b},
Sophie Merlin^{a,b}, Christophe Audebert^{a,b}

a Gènes Diffusion, 3595, Route de Tournai, 59501 Douai, France

b PEGASE-Biosciences, Institut Pasteur de Lille, 1 Rue du Professeur Calmette, 59019 Lille, France

1. Introduction

Artificial insemination (AI) is a common assisted reproductive technique used in the swine industry that consists in mechanically introducing diluted semen containing viable spermatozoa into the female genital tract. Unlike other animal productions, pig semen is commonly used diluted, in liquid-stored at a relatively controlled temperature (between 14°C - 18°C). This method requires a rather rapid use of the semen (< 5 days). A main challenge for the artificial insemination industry is to maintain semen quality that satisfies the fertilization objectives while preserving bacteriological and virological safety. A relatively large number of studies have identified bacteria recovered from boar semen attributed to fecal, preputial, cutaneous microorganisms, or other environmental bacteria from semen analysis and semen dose packaging areas for artificial insemination¹. From automated semen collection technologies² to the use of antibiotics incorporated at the time of semen dilution, many procedures have made it possible to bacteriologically secure the product³. Several bacteria of the Enterobacteriaceae family have been associated with significant decreases in sperm motility⁴, *E. coli* have been associated with decreases in litter size⁵, *P. aeruginosa* alters sperm capacitation *in vitro*⁶. Bacteriospermia is a key risk factor which should be controlled to avoid product alterations that would deteriorate the expected fertility performance¹. In addition to the objective of quality and safety, the reduction of the antibiotic load in marketed semen is an objective of the pig semen industry⁷. Basic culture methods do not allow the complexity of a microbiota to be revealed⁸ and in a production context, this type of method can be difficult to implement for practical reasons, response time and/or cost-effectiveness. For more than ten years, with the arrival of high-throughput DNA sequencers, microbiota investigation approaches have considerably evolved, initially reserved for research in human health or studies of microbial ecosystems in ecology, today these methods are beginning to be used in an industrial context to qualify products in bacteriological terms⁹. The study of the human seminal microbiota was performed using 16S high throughput sequencing¹⁰ and microbial signature has been associated with semen quality. The proportion of *Lactobacillus* and *Gardnerella* was significantly higher in normal samples, while *Prevotella* was significantly higher in low quality samples. Seminal microbiota is not well-known, Weng *et al.* attempted in their study to characterize bacterial communities in the human sperm of infertile couples. Their results of the analysis showed that the types of seminal bacterial communities were strongly associated with sperm health¹⁰.

In this study, we offer a preliminary approach integrating the high-throughput metagenomics sequencing tool into the study of porcine bacteriospermia. We will attempt to characterize the porcine seminal microbiota before and after dilution (at the beginning and at the end of the process) of different semen production artificial insemination center: boar studs (AI) (5 French AI studs and 1 AI stud in the United States of America). This pilot study will demonstrate the potential application of seminal microbiota analysis in the context of industrial boar semen production. It consists in an experimental design and methodology to assess the use of this innovative approach. The interpretation of the results allows to perceive which variables would be preferentially associated with the composition of the seminal microbiota.

2. Material and methods

2.1. Sample collection

In the 6 participating boar studs, 5 French studs (FRSF1, FRSF2, FRSF3 respectively for French Slatted Floor 1, 2 and 3 and FRSD1 et FRSD2 for French Saw Dust 1 and 2) and 1 US stud (USSF1 for USA Slatted Floor 1), a minimum of 5 boars were collected 3 times over 3 weeks. For each boar sample, a set of variables was collected in order to compare them with the composition and microbial diversity: its center and country of belonging, the type of soil, the presence of air filtration in the center, the size of animal housing, its genetic lineage and its age at collection (in days) (see Table 1). Each ejaculate was split into two samples, one before dilution ("raw" semen), the other after dilution and industrial processing. Each sample was extended in Elixir (GenePro Inc, Madison WI). Elixir contains the following classes of antibiotics: aminoglycoside, β -lactam, Aminocyclitol and fluoroquinolone antibiotic. Immediately after collection or processing the samples are frozen in sterile tubes 1.5 mL at -20°C. The selected samples for the study did not show any problems of non-compliance regarding reproductive performance (comparable to the average observed total born piglets per parity). All boar studs included in the study belong to Gènes Diffusion company, so they have relatively similar collection practices, identical extenders and an identical quality management system.

Table 1: sample description and associated variables

2.2. DNA extraction

Each sample was thawed and homogenized by vortex, and DNA was extracted with Nucleospin® 96 Soil kit (Macherey Nagel) from the sample a volume of 200 μ L (as a biological matrix) placed in the tubes containing the supplied beads. Under aseptic conditions, the extraction was performed by adding SL2 and SX buffers in accordance with the supplier's introductions. The tubes then are subjected to intense 30 Hz agitation for 2 min with beadmills (Retsch, Haan). The protocol followed the supplier's recommendations until the elution phase that occurred after the incubation of the plates with 75 μ L of TE 1X preheated at 70°C followed by a 1-hour incubation at room temperature and finally an elution phase after centrifugation at 6000 x g for 2 min.

2.3 High-throughput sequencing

2.3.1 Library preparation

The library is based on the principle of dual-indexing thanks to paired-end sequencing. To achieve this, two PCRs are successively performed: From 2 μ L of the extracted DNA, a first PCR in final volume of 50 μ L, MgSO₄ 1mM, 2,5 U of Precision Taq Polymerase (Applied Biological Materials Inc, Richmond), each primer has a final concentration of 500 nM. For this first PCR, forward and reverse primers have been designed with a 5'-Tag sequence respectively TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG as a forward primer, GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG as ther reverse primer and a 16S ARNr gene specific sequence respectively CCTACGGGNGGCWGCAG as the forward primer and GACTACHVGGGTATCTAATCC as the reverse. According to *E. coli* 16S ARNr sequence gene specific primers target a locus between position 341 and 785, resulting in the amplification of a locus of 445 pb. The amplification conditions were 3 min at 94°C, 25 cycles of 15 s at 94°C for denaturation, 15 s at

51°C for primers annealing and 45 s at 68°C for extension, followed by an incubation at 68 °C for 1 min. At end of this first PCR, amplification products have been purified with NucleoFast® 96 PCR (Macherey Nagel) according to supplier recommendations except for the last step for which 35 µL of TE 1X preheated at 70°C were used for elution. The purified products constitute the matrix of the second PCR. From 5 µL of the previously purified DNA, a second PCR was performed in final volume of 50 µL, MgSO4 1mM, 2,5 U of Precision Taq Polymerase (Applied Biological Materials Inc, Richmond). Each primer has a final concentration of 500 nM. The amplification conditions were the same as those of the previous one except the number of cycles reduced to 8. In addition to the Tag sequences, these PCR2 primers contain a locus to index the samples (barcode sequence) and a locus sequence adapter suitable for the Illumina sequencing technology. A Nucleospin® purification step identical to the one presented above is performed followed by a Quant-iT PicoGreen ds DNA quantification (Life Technologies). An equimolar pool of index is produced, 100 µL of this mixture is purified by Nucleomag® (Macherey Nagel). The purification is performed twice with 140 µL of beads suspension at each purification to conclude with a final elution in 50 µL of TE 1X. This purified mixture is diluted 10 and 20 times and is then monitored by Bioanalyzer in high sensitivity mode (Agilent).

2.3.2 Sequencing

Sequencing library has been paired-end sequenced on Miseq platform with MiSeq Reagent Kit v3 allowing 600 sequencing cycles to be performed. At the end of sequencing, a quality control by FastQC was carried out.

2.4. Metagenomics analysis

2.4.1. From sequencing to microbiota composition

The 16S rRNA gene libraries were sequenced with 2 × 300 bp paired-end reads on the MiSeq sequencer (Illumina), using MiSeq v3 illumina reagent kits. Quality of the raw sequence data was checked with the FastQC quality-control tool, and the datasets were analyzed with QIIME 1.9 pipeline using the GreenGenes database version 13.08. Firstly, paired-end sequence reads were merged together using PEAR software, forming one sample from each two replicates. Sequence reads were then filtered with a quality-score acceptance rate of 20 or better and a minimum and maximum size of 156 and 517 respectively. Chimeric sequences were removed with VSEARCH. OTU (Operational Taxonomic Unit) clustering was performed by QIIME, using an open reference picking pipeline with SortMeRNA (for reference picking) and sumacust (for *de novo* OTU picking) generating an operational taxonomic unit (OTU) table. The OTU counts were normalized using Cumulative Sum Scaling (CSS) algorithm in order to minimize the effect of inter-sample variation like library size difference.

2.4.2. Statistical Analyses

Statistical analyses of the 16S rRNA gene sequence data were performed with QIIME statistical tools and R language. All analyses were made from the two OTU tables, normalized and non-normalized. To study the bacterial diversity of the samples, α -diversity metrics were computed on the non-normalized table, and α -rarefaction plots were generated with QIIME α -rarefaction command. Statistically significant differences in the α -diversity (*i.e.*, in average Shannon index values, chao1 values and PDwholeTree values) were computed using R stats packages, applying non-parametric methods and considering p-value < 0.05 as statistically significant. To analyze the differences in the overall bacterial diversity (β -diversity) across the samples, weighted UniFrac distance matrices were generated from the normalized OTU table, and principal coordinate analysis (PCoA) plots were produced. These PCoA plots were visualized with the EMPeror data visualization program. To ascertain whether the visually

observed differences were statistically significant, normalized OTU table was exported into STAMP software, a package to analyze taxonomic profiles that promotes “best practices” in choosing appropriate statistical techniques and reporting results.

3. Results

At the end of the sequencing run and before the assembly step, we obtained a total of 7,983,458 sequences from the 192 paired samples. During the assembly step of the paired reads, 94.52% of the reads were assembled, the rejected reads being considered impossible to assemble (5.2%) or were simply discarded (0.2%) (See Table 2). During the quality step, 2.4% of the reads are removed, not meeting the minimum quality (phred-score) or size criteria. 7,423,334 sequences are thus retained before the filter step of the chimeric sequences (See Table 2). 396,837 reads are removed, considered as uncertain (1.7% of deleted reads) or detected as chimeric reads (3.7% of deleted reads). At the end of this step 7,026,497 sequences passed all the quality filters before the OTU clustering step. At the end of this OTU clustering step, 1 outlier individual is removed from our samples, the number of sequences (7019) being well below the average (36556). To summarize, 99.5% of the indexes and 87.9% of the initial reads are considered relevant for the analysis and data interpretation.

Table 2: metrics related to the performed sequencing run

3.1. Which variables are preferentially associated with microbiota profiles?

The objective of the study was to identify the environmental variables associated with a variation in microbial composition and diversity. The variables can be dissociated as follows: boar-related variables, such as its genetic lineage (Pietrain / Landrace / Duroc) and the age of the boar at the time of collection. A second group of variables is related to the boar's environment: the country (USA or France) and its associated stud (1 stud in USA and 5 studs in France), the type of flooring (sawdust or slatted floor) as well as the presence of an air filtration system in the building (binary variable).

The genetic lines and age of the boar collected can potentially have a direct or indirect impact (due to the length of the animal's hair, for example) on the composition of the seminal microbiota. The experimental design of this pilot study does not allow to elucidate the associated variables and explain the variability of the observed seminal microbial profiles. Even if the correlation between age and microbial diversity seems to be negative, this is not statistically verified (Kendall correlation – p-value > 0.46). Indeed, several variables can be qualified as confounding (see Table 1), such as size of animal housing and floor type. However, it seems that the profile of the seminal microbiota can be associated with the stud producing the semen.

It appears that the samples from the American stud have a composition that is difficult to compare with those observed in the French studs. Indeed, for microbiota observations obtained before dilution, a bacterial community showing a higher α -diversity (Fig. 1b) and a family of Ruminococcaceae (pink bar on Fig. 1a) is more present than in samples from French studs. For French studs, the families of Pseudomonadaceae and Enterobacteriaceae (respectively orange and green bars on Fig. 1a) are predominant compared to the American stud. As shown in Figure 1, the seminal microbiota profiles in the USA are relatively different -both in terms of composition (Fig. 1a) and α -diversity (Fig. 1b)- from those observed in French studs. Multiple factors, from factors related to the American stud of the study, to factors related to the international transport of samples, could explain this microbiota profile. In the rest of the study, samples from the American stud were not statistically processed to avoid erroneous analytical use of the confounding variables.

Figure 1: a) Bar plot representing the proportions of the different bacterial families constituting porcine seminal microbiota in raw semen samples (before dilution) according to the different boar studs of the study; b) Box plots showing α -diversity (Chao1 richness estimator).

Despite homogenous semen collection and processing in the French studs studied, differences in the composition of the α -diversity index can be observed before semen processing (before extending and packaging).

At the taxonomic order level, for slatted floors Clostridiales are significantly more abundant 15.10% (3.40% for sawdust), Lactobacillales 7.80% (2.50% for sawdust), Bacteriodales 6.10% (1.80% for sawdust). Whilst for sawdust, two Gram negative bacterial orders are significantly more abundant: Enterobacteriales 17.90% (7.80% for slatted floors) and Rhizobiales 11.80% (8.40% for slatted floors). Rhizobiales, consists of nitrogen-fixing and symbiotic bacteria in plant roots, sawdust could be more favorable for their development (Fig. 2)

Figure 2: Significant differences (q -value < 0.05) are represented here between the two groups of microbiota in orange for sawdust and in blue for slatted floors.

3.2. The flooring may have a significant impact on the seminal microbiota

As shown in the figure 2, where the different seminal microbiota of the samples are projected onto a Principal Component Analysis (PCA), taking into consideration their β -diversity, it appears that two groups be identified according to the floor type. A seminal microbiota found in the raw semen seems to be associated with sawdust, another relatively different one is associated with slatted floors. In addition, the α -diversity indices show significantly less diversity for the FRSD2 and FRSD1 studs, both with sawdust as floor type, compared to the FRSF2, FRSF3 and FRSF1 β -diversity indices with slatted floors (Fig. 3).

Figure 3: before dilution microbiota compound projection PCA (β -diversity), red plot for slatted floor, blue plot for sawdust

No other association with the seminal microbiota seems obvious with other variables, so no association has been found with the use of filtered or unfiltered air, with the genetic line of the animals, the country of origin of the animals (the sizes of the studs in this study cannot be completely dissociated from the floor type, indeed studs using sawdust clearly host the largest number of animals).

3.3. The process impact on the microbial diversity of the seminal microbiota

Before dilution, the α -diversity of the porcine seminal microbiota is significantly different between the French studs considered in this study (Fig. 4, p -value < 0.001). On the other hand, after dilution, this difference no longer appears significant (Fig. 5, p -value > 0.05).

Figure 4: Box plots of observed OTUs richness index of raw semen distinguishing between French boar studs

As noted by the analysis of α -diversity before and after dilution (Fig. 4 and 5), it seems that the differences in the value of the diversity indices are lower after dilution. In addition, after pairing the samples, it is possible to show that, on the α -values, three effects can be distinguished:

- FRSD1 using sawdust which shows the lowest diversity value before dilution, shows a slight increase in this index, after (p-value = 0.03) with a small standard deviation of measurement

- FRSF3 and FRSF2, two studs using slatted floors which show the highest diversity value before dilution, show a very significant decrease (p-value of 0.0022 and 0.005 respectively)

- FRSF1 using slatted floors and FRSD2 using sawdust both studs show an insignificant downward trend

If, depending on the boar stud, a large number of bacterial families differ before dilution (mainly due to the type of flooring) after dilution, the Pseudomonadaceae family proportion is the only significant difference due to a single stud (Lactobacillaceae, Enterococcaceae, Clostridiaceae show a significant p-value with little proportion lower than 1.5%).

Figure 5: Boxplots of observed OTUs richness index of diluted semen distinguishing between French boar studs

3.4. 16S metagenomics as a putative tool to ensure the conformity of a semen production process

Mean relative abundance of Pseudomonadaceae is 5.10% for slatted floors and 6.60% for sawdust in raw seminal microbiota. Pseudomonadaceae are commensal of the pig nasal passages and saprophytes of hydric environment. Elevated proportion of Pseudomonadaceae may indicate the development of a bacterial layer on surfaces within a distribution system (Figure 6, p-value < 0.001).

Figure 6: Boxplot displaying the relative abundance of the family of Pseudomonadaceae according to AI studs

While most studs, after dilution, show a proportion of Pseudomonadaceae relatively similar to that observed on raw semen, it appears that the proportion of Pseudomonadaceae at the FRSF3 stud is at least twice as high than the other studs with extreme samples showing large proportions of Pseudomonadaceae.

4. Discussions

4.1. Dilution process contributes to the standardization of production

α -diversity comparison before and after processing (extending semen) highlights a reduction of α -diversity after extending process. Beyond this decrease, it appears that the amplitude of the differences between samples from different boar studs tend to fade. Thus, dilution would make it possible to standardize semen production by diluting the bacterial concentration (obviously the

addition of antibiotics would make it possible to control bacterial proliferation) and erasing the differences that can be observed between the centers before dilution.

4.2. A tool to monitor bacteriospermia

Cultivation methods are both unlikely to be feasible in terms of cost and practicality, and they do not reveal the complexity of a relatively complex microbiota. Whereas molecular methods for portion sequencing of the 16S gene allow relatively fast access to a vision that is not perfect yet sufficient to compare a standard production sample with a produced batch, for example. This approach makes it possible, when producing living biological material, to identify bacterial communities that are likely to alter the product or its performance. In regards to semen this could explain certain fertility issues. A recent study of the human seminal microbiota suggests that a small subset of bacteria could affect sperm physiology during transition, more specifically motility and morphology¹¹. These suggestions tend to support the fact that the metagenomic tool should be used, in the relatively near future, as a tool to monitor the bacterial microbiota. Even if these samples did not show any compliance problems on the reproductive performance level, the hypothesis that a biofilm had developed within the semen dilution and packaging system is possible in FRSF3 stud. This result shows in a commercial production context that semen can be monitored in terms of microbiota to ensure compliance or to anticipate biofilm development problems. A microflora shift in the product can be an early indicator of a change in the production system.

4.3. A perspective to adjust the dilution conditions of the semen

Reducing the use of antibiotics in livestock farming is a priority for the animal industry, particularly in France (EcoAntibio plan). Of course, the presence of antibiotics has the following objectives: to kill contaminating bacteria before they have an opportunity to grow or challenge for nutrients in the sperm extender, the bacteria are not transmitted to the female by AI (unless they are resistant). The presence of antibiotics in a semen extender may have more unwanted effects: antibiotics may be toxic to spermatozoa, act as a selection pressure for resistant bacteria or can release LPS. Our results on porcine bacteriospermia show, for example, that in some cases dilution with antibiotics may favor the emergence of taxa that are not very sensitive to these antibiotics, which are therefore found in a higher proportion after dilution. This was observed on the FRSF3 stud for the taxon Pseudomonadaceae

The monitoring of bacterial communities linked to production sites could make it possible to adapt the antibiotic mix used in boar studs. Today, a small number of extender/antibiotic formulas exist to cover a large number of studs with possibly different seminal microbiota. In the relatively near future, it is conceivable, depending on the measurement of seminal microbiota by boar stud, to adapt the antibiotic formulas present in the extenders according to these measurements. Promising methods such as centrifugation using a low-density colloid method can reduce the overall bacterial load¹². This adaptation could reduce the amount of antibiotics used without increasing the risk. These alternative methods could be evaluated with metagenomic methods.

4.4. The main limitations of the metagenomics method

The metagenomic approach makes it possible to evaluate the complexity of a given microbiota without a real equivalent method to date. Nevertheless several elements may constitute a limit of application or require the use of additional analyses. Thus, the overall bacterial load (quantitative analysis) is not feasible by metagenomics, an evaluation could be carried out by qPCR. In addition, metagenomic analyses may often not allow access to the taxonomic level of the bacterial species which does not allow to conclude categorically on the potential pathogenicity of a sample. As a matter of principle and

because of this modality, metagenomics being based on a portion of the 16S gene, taxonomic sensitivity may sometimes not allow to conclude on important events such as the presence of a pathogenic element.

5. Conclusion

This pilot study shows that, as it has been shown in humans before, the porcine seminal microbiota may have an impact to consider in the near future. Our study focused on the production parameters that are found to have an impact on the microbial composition of porcine seminal fluid. Thus, we were able to observe that the boar stud influenced the composition of this microbiota. It appears that one of the variables most associated with this composition is the type of flooring on which the animal is housed. In the future, we plan to establish reference samples (standards / stud) for which the semen quality control parameters (CASA- computer assisted sperm analysis) as well as fertility performance in commercial farms. The samples can then be compared in terms of seminal microbiota. This monitoring will make it possible to bacteriologically control the products leaving the stud and possibly highlight process problems or problems inherent to an individual. Thus, it seems particularly interesting to monitor the samples after dilution. Even if these samples may appear less "different" bacteriologically, this approach may allow the identification of dilution problems and the adaptation of the antibiotic mix with greater precision.

Acknowledgments

The authors would like to thank all the sampling technicians for having carried out the sampling according to our recommendations: Jody Frankum, Patricia Murphy, Lise Wright, Olivier Nourry, Valérie Radenac, Mathieu Lebel, Marie Ouadah, Elodie Curial, André Houdin, Béatrice Behetre, Arnaud Boscher, Philippe Bellenguez, Alain Didier, Frédéric Cariou, Fabien Cevaër, Julie Clahèvre, Jose Le Caroff and Stéphane Abgrall.

1.Kuster, C. E. & Althouse, G. C. The impact of bacteriospermia on boar sperm storage and reproductive performance. *Theriogenology* **85**, 21–26 (2016).

2.Aneas, S. B., Gary, B. G. & Bouvier, B. P. Collectis 1 automated boar collection technology. *Theriogenology* **70**, 1368–1373 (2008).

3.Rodriguez, A. L., Soom, A. Van, Arsenakis, I. & Maes, D. Boar management and semen handling factors affect the quality of boar extended semen. *Porcine Health Manag* 1–12 (2017).

4.Úbeda, J. L. et al. Adverse effects of members of the Enterobacteriaceae family on boar sperm quality. *Theriogenology* **80**, 565–570 (2013).

5.Maroto Martín, L. O. et al. Bacterial contamination of boar semen affects the litter size. *Anim. Reprod. Sci.* **120**, 95–104 (2010).

6. Sepúlveda, L., Bussalleu, E., Yeste, M. & Bonet, S. Effect of *Pseudomonas aeruginosa* on sperm capacitation and protein phosphorylation of boar spermatozoa. *Theriogenology* **85**, 1421–1431 (2016).

7.Schulze, M. Cationic Synthetic Peptides : Assessment of Their Antimicrobial Potency in Liquid Preserved Boar Semen. *PLoS One* **9**, 1–8 (2014).

338 8.Amann, R. I., Ludwig, W. & Schleifer, K. Phylogenetic Identification and In Situ Detection of Individual
339 Microbial Cells without Cultivation. *Microbiol Rev* **59**, 143–169 (1995).

340 9.Siegwald, L., Audebert, C., Even, G. & Viscogliosi, E. Data Descriptor : Targeted metagenomic
341 sequencing data of human gut microbiota associated with *Blastocystis* colonization. *Sci Data* **346** 1–7
342 (2017).

343 10.Weng, S. L. et al. Bacterial communities in semen from men of infertile couples: Metagenomic
344 sequencing reveals relationships of seminal microbiota to semen quality. *PLoS One* **9**, (2014).

345 11.Baud, D., Pattaroni, C., Vulliemoz, N. & Castella, V. Sperm Microbiota and Its Impact on Semen
346 Parameters. *Front Microbiol.* **10**, 1–9 (2019).

347 12. Morrell JM, Núñez-González A. et al. Removal of bacteria from boar semen using a low-density
348 colloid Theriogenology **126**, 272-278 (2019).

349

350 Tables & figures

351 Table 1: sample description and associated variables

352

AI center id	Country	floor type	number of sampled boars	filtered air	Size of animal housing	sampled genetic strain	average age of collected boars (days)	standard deviation of age (days)
FRSF1	France	slatted floors	5	no	129	Pietrain	474.13	125.62
FRSF2	France	slatted floors	5	yes	140	Pietrain	603.25	315.85
FRSF3	France	slatted floors	5	yes	121	Landrace	458.38	126.98
USSF1	USA	slatted floors	6	no	150	Duroc	536.75	95.37
FRSD1	France	saw dust	5	no	435	Pietrain	770.00	285.97
FRSD2	France	saw dust	6	no	367	Pietrain	798.63	481.42

353

354

355

356
357

Table 2: metrics related to the performed sequencing run

STEP	RESULTS
Total number of reads after sequencing	7,983,458
Number of reads after assembly	7,605,999
Number of reads after quality filtering	7,423,334
Number of reads after chimeric detection	7,026,497
Mean reads [median] by index	38663 [30498]
Min - Max reads by index	7252 - 228558
Number of outliers (< 10K reads)	1
Final percentage of meaningful index	99.5%
Final percentage of meaningful reads	87.9%

358

359

360

361 Figure 1: a) Barplot representing the proportions of the different bacterial families constituting porcine
362 seminal microbiota in raw semen samples (before dilution) according to the different boar studs of
363 the study; b) Box plots showing α -diversity (Chao1 richness estimator).
364

Figure 2: pre-dilution microbiota compound projection PCA (β -diversity), red plot for slatted floor, blue plot for sawdust

Figure 3: Significant differences (q -value < 0.05) are represented here between the two groups of microbiota in orange for sawdust and in blue for slatted floor

Figure 4: Boxplots of observed OTUs richness index of raw semen between French boar studs

Figure 5: Boxplots of observed OTUs richness index of diluted semen between French boar studs

Figure 6: Boxplot displaying the relative abundance of the family of Pseudomonadaceae according to

389
390

the boar stud

391
392
393
394