Title: Characterisation and comparison of semen microbiota and bacterial load in men with infertility, recurrent miscarriage, or proven fertility

Running title: Characterisation and comparison of semen microbiota and bacterial load in men with infertility, recurrent miscarriage, or proven fertility

Authors: Shahriar Mowla[1] , Linda Farahani[2,3] , Tharu Tharakan[3,4], Rhianna Davies[3], Gonçalo D S Correia[1,5], Yun S Lee[1,5], Samit Kundu[1,5], Shirin Khanjani[6], Emad Sindi[3], Raj Rai[2], Lesley Regan[2,7], Dalia Khalifa[8], Ralf Henkel[3,9,10], Suks Minhas[4], Waljit S Dhillo[3], Jara Ben Nagi[11], Phillip R Bennett[1,5,7], David A MacIntyre[1,5,7#], Channa N Jayasena[3,8#*]

1. Institute of Reproductive and Developmental Biology, Imperial College London, UK.

2. Wolfson Fertility Unit, Department of Gynaecology, St. Mary's Hospital, Imperial College NHS Trust, UK.

3. Section of Endocrinology & Investigative Medicine, Imperial College London, UK.

4. Department of Urology, Charing Cross Hospital, Imperial College NHS Trust, UK.

5. March of Dimes European Prematurity Research Centre, Imperial College London, UK.

6. Department of Gynaecology, University College London Hospital, UK.

7. Tommy's National Centre for Miscarriage Research, Imperial College London, London, UK

8. Department of Andrology, Hammersmith Hospital, Imperial College NHS Trust.

9. LogixX Pharma, Theale, Berkshire, UK

10. Department of Medical Bioscience, University of the Western Cape, Bellville, South Africa

11. Centre for Reproductive and Genetic Health (CRGH), London, UK.

#These authors contributed equally.

* Channa Jayasena. 6th Floor Commonwealth building, Imperial College Faculty of Medicine, Hammersmith Hospital, Du Cane Rd, London. W12 0HS c.jayasena@imperial.ac.uk

Abstract:

Several studies have associated seminal microbiota abnormalities with male infertility but have yielded differing results owing to their limited sizes or depths of analyses. The semen microbiota during recurrent pregnancy loss (RPL) has not been investigated. Comprehensively assessing the seminal microbiota in men with reproductive disorders could elucidate its potential role in clinical management. We used semen analysis, terminal-deoxynucleotidyl-transferase-mediated-deoxyuridine-triphosphate-nick-end-labelling, Comet DNA fragmentation, luminol ROS chemiluminescence and metataxonomic profiling of semen microbiota by16S rRNA amplicon sequencing in this prospective, cross-section study to investigate composition and bacterial load of seminal bacterial genera and species, semen parameters, reactive oxidative species (ROS), and sperm DNA fragmentation in men with reproductive disorders and proven fathers. 223 men were enrolled included healthy men with proven paternity (n=63), the male partners in a couple encountering RPL (n=46), men with male factor infertility (n=58), and the male partners of couples unexplained infertility (n=56). Rates of high sperm DNA fragmentation, elevated ROS and oligospermia were more prevalent in the study group compared with control. In all groups, semen microbiota clustered into three major *genera*-dominant groups (1, Streptococcus; 2, Prevotella; 3, Lactobacillus and Gardnerella); no species clusters were identified. Group 2 had the highest microbial richness (P<0.001), alpha-diversity (P<0.001), and bacterial load (P<0.0001). Overall bacterial composition or load has not found to associate with semen analysis, ROS or DNA fragmentation. Whilst, global perturbation of the seminal microbiota is not associated with male reproductive disorders, men with unidentified seminal *Flavobacterium* are more likely to have abnormal seminal analysis. Future studies may elucidate if *Flavobacterium* reduction has therapeutic potential.

Keywords: Semen, microbiota, male infertility, genitourinary infection, recurrent pregnancy loss

Main text:

Introduction

Mean sperm counts reported within clinical studies have reduced annually by 2.6% since 2000 (1). Male factor accounts for approximately half of all cases of infertility yet there are limited available interventions to improve sperm quality. Understanding the pathogenesis of male infertility may reveal novel therapeutic approaches for treating affected couples.

Symptomatic, genitourinary infection is an established cause of male infertility which may be detected by semen culture and treated with antibiotics (2, 3). The current European Association of Urology guidance states that whilst antibiotics may improve overall semen quality, there is no evidence of increased pregnancy rates after antibiotic treatment of the male partner (4) (5) . Seminal leukocytes release bactericidal reactive oxygen species (ROS) in response to infection, however paradoxically this may damage sperm DNA and impair semen quality (6). We and others have reported that asymptomatic men affected by recurrent pregnancy loss (RPL), infertility or impaired preimplantation embryo development have increased risks of high seminal ROS and sperm DNA fragmentation (7, 8, 9, 10, 11, 12). It is therefore plausible that asymptomatic seminal infections may be associated with impaired reproductive function in some men. Since semen culture has a limited scope for studying the seminal microbiota due to its inability to identify all present microbiota next generation sequencing (NGS) approaches have been reported recently by a growing number of investigators (13, 14, 15, 16, 17, 18, 19). These studies, with varying methodologies, have produced inconsistent and conflicting results. As such, the composition of semen microbiota and its associations with clinical and molecular markers of male reproduction remains understudied. Elucidation of an association would have wide clinical application with therapeutic potential couple with reproductive disorders (20).

We hypothesised that semen microbiota composition associates with functional semen parameters, including ROS levels and sperm DNA fragmentation. To test this, we explored relationships between metataxonomic

profiles of bacteria, bacterial copy number and key parameters of sperm function and quality in semen samples collected from 223 men, including those diagnosed with male factor infertility, unexplained infertility, partners affected by recurrent miscarriage, and paternity-proven controls.

Methods

Ethical approval was granted by the West London and Gene Therapy Advisory Committee (GTAC) Research Ethics Committee (14/LO/1038) and by the Internal Review Board at the Centre for Reproductive and Genetic Health (CRGH) (IRB-0003C07.10.19). Participants were recruited following informed consent from clinics in Imperial College London NHS Trust and The Centre for Reproductive and Genetic Health (CRGH). Further detailed information on methods used in this study are included in the Supplementary Material.

Semen samples were produced by means of masturbation after 3-7 days abstinence. All semen samples were collected into sterile containers after cleaning of the penis using a sterile wipe. Samples were incubated at 37°C for a minimum of 20 mins prior to analysis. An aliquot was collected in a sterile cryovial and stored at -80°C.

Diagnostic semen analysis was carried out according to WHO 2010 guidelines and UK NEQAS accreditation (21) (22). Seminal analysis was performed in the Andrology Departments of Hammersmith Hospital and CRGH. Microscopic and macroscopic semen qualities were assessed within 60 mins of sample production. Semen volume, sperm concentration, total sperm count, progressive motility and total motility count, morphological assessment, anti-sperm antibodies and leucocyte count were established.

ROS analysis was performed using an in-house developed chemiluminescence assay validated by Vessey et al (23). Results are therefore reported as 'relative light units per second per million sperm'. The upper limit of optimal ROS was internally determined at 3.77 RLU/sec/10^6 sperm (95% CI) (24).

Sperm DNA fragmentation assessment performed by TUNEL (Terminal deoxynucleotidyl transferase biotin-dUTP Nick End Labelling) assay defined elevated sperm DNA fragmentation as >20% (25). Samples for the

COMET assay were sent to the Examen Lab (Belfast, UK) for analysis with elevated sperm DNA fragmentation defined as >27% (26).

DNA extraction was performed on 200 µL of semen using enzymatic lysis and mechanical disruption. Bacterial load was estimated by determining the total number of 16S rRNA gene copies per sample using the BactQuant assay (27).

Metataxonomic profiling of semen microbiota was performed using MiSeq sequencing of bacterial V1-V2 hypervariable regions of 16S rRNA gene amplicons 16S rRNA genes using a mixed forward primerset 28F-YM GAGTTTGATYMTGGCTCAG, 28F-Borrelia GAGTTTGATCCTGGCTTAG, 28F-Chloroflex GAATTTGATCTTGGTTCAG and 28F-Bifdo GGGTTCGATTCTGGCTCAG at a ratio of 4:1:1:1 with 388R reverse primers. Sequencing was performed on the Illumina MiSeq platform (Illumina, Inc. San Diego, California). Following primer trimming and assessment of read quality, amplicon sequence variants (ASV) counts per sample were calculated and denoised using the Qiime2 pipeline (28) and the DADA2 algorithm (29). ASVs were taxonomically classified to species level using a naive Bayes classifier trained on all sequences from the V1-V2 region of the bacterial 16S rRNA gene present in the SILVA reference database (release 138.1) (30) (31).

Further methodological detail can be found in supplementary text.

Controls and contamination 3 negative kit/environmental control swabs were included to identify and eliminate potential sources of contamination and false positives in the 16S m*etataxonomic profiles.* These swabs were removed from the manufacturers packaging, waved in air, and then subjected to the same entire DNA extraction protocol. Decontamination of data was done using the decontam package (v1.9.0) in R, at ASV level, using both “frequency” and “prevalence” contaminant identification methods with *threshold* set to 0.1 (31). The “frequency” filter was applied using the total 16S rRNA gene copies measured as the *conc*

parameter. For the “prevalence” filter all 3 blank swabs were used as negative controls and compared against all semen samples. ASVs classified as a contaminant by either method (n = 94) were excluded.

Statistical analysis. Hierarchical clustering with Ward-linkage and Jensen-Shannon distance was used to assign samples to putative community state types, with the number of clusters chosen to maximise the mean silhouette score. Linear regression models used to regress microbiota features against semen quality parameters and other clinical and demographic variables were fitted with the base R *lm* function (v4.2.0). The Benjamini-Hochberg false discovery rate (FDR) correction was used to control the FDR of each covariate signature independently (e.g., ROS, DNA Fragmentation, or Semen quality), with a $q < 0.05$, or 5%, cut-off, in both regression and Chi-squared analyses. Detailed information for statistical modelling is presented Supplementary methods.

Results

Study population: Semen samples were collected from a total of 223 men; this included control (n=63) and a study group (n=160) comprised of men diagnosed with male factor infertility (MFI) (n=58), male partners of women with recurrent pregnancy loss (RPL) (n=46) and male partners of couples diagnosed with unexplained infertility (UI) (n=26). The overall mean age of the total cohort was 38.1 ± 6 (mean ± SD). The mean age for controls was 40.1 ± 8, and the mean age for patients undergoing various fertility investigations was 37 ± 4.8. Ethnicity representation amongst recruited cohorts were not significantly different (p=0.38, Chi-square; Supplementary Table 1).

Semen quality assessment: Rates of high sperm DNA fragmentation, elevated ROS and oligospermia were more prevalent in the study group compared with control (Table 1). The study group represented 85% of samples with high sperm DNA fragmentation, 85% of samples with elevated ROS and 79% of samples with oligospermia. Rates of abnormal seminal parameters including low sperm concentration, reduced progressive motility and ROS concentrations were found to be highest in the MFI group (Supplementary Figure 1). Baseline characteristics between the RPL, unexplained subfertility and controls groups were similar (Supplementary Figure 1)

Seminal microbiota: Following decontamination, a total of 7,998,565 high quality sequencing reads were identified and analysed. Hierarchical clustering (Ward linkage) of relative abundance data resolved to genera level identified three major clusters, as determined by average silhouette score, amongst all samples (Figure 1, Supplementary Figure 2). These were compositionally characterised by highest mean relative abundances of 1. *Streptococcus* (23.8%), 2. *Prevotella* (24.4%), or 3. *Lactobacillus* and *Gardnerella* (35.9% and 35.6%, respectively, Figure 1C). Assessment of bacterial load using qPCR showed Clusters 2 and 3 had significantly higher bacterial loads compared to Cluster 1. Similar analyses were performed using sequencing data mapped to species level, however, examination of individual sample Silhouette scores within resulting clusters highlighted poor fitting indicating a lack of robust species-specific clusters

(Supplementary Figure 3). To further investigate potential pairwise ecological interactions between taxa, a co-occurrence analysis was performed on the sequencing data, mapped to species level, with the SparCC algorithm (Figure 2). 5 major graph communities were detected. Community 1 highlighted a co-occurrence pattern between *Gardnerella vaginalis* and *Lactobacillus iners*, in agreement with the composition of cluster 3 from the hierarchical clustering analysis at genera level. Taxa belonging to communities 3 and 4 had a high number of connections (higher node degree), both within and between the two communities, including some anti-co-occurrence patterns (SparCC $\rho < 0$). These communities included species from Genera *Staphylococcus, Peptoniphilus, Corynebacterium, Prevotella,* among others.

Bacterial richness, diversity and load were similar between all patient groups examined in the study (Supplementary Figure 4). Similarly, no significant associations between bacterial clusters, richness, diversity or load with seminal parameters, sperm DNA fragmentation or semen ROS were observed (Supplementary Tables 2-3). No significant differences in relative abundance of bacterial taxa between patient groups were detected, at genus or species-level. Several organisms at the genus level, identified variably in the literature as responsible for genito-urinary infection, were observed in our dataset but their prevalence did not reach our criteria (present in at last 25% of the samples) to be carried forward to regression modelling (32) (33) (24). This included *Chlamydia, Ureaplasma, Neisseria, Mycoplasma* and *Escherichia*. However, several associations ($p<0.05$) between relative abundance of specific bacterial genera and key sperm parameters were observed (Table 2). These included increased sperm DNA fragmentation was positively associated with increased relative abundance of *Porphyromonas* and *Varibaculum* and inversely correlated with *Cutibacterium* and *Finegoldia*. ROS was positively associated with *Lactobacillus* species relative abundance, with analyses performed at species level taxonomy indicating that this relationship was largely driven by *L. iners* ($p=0.04$; Table 3). In contrast, *Corynebacterium* was inversely associated with ROS and positively associated with semen volume. Of note, *Flavobacterium* genus was positively associated with both abnormal semen quality and sperm morphology and in both cases, withstood FDR correction for multiple testing ($q=0.02$ and $q=0.01$, respectively) (Table 2) (Figure 3). Consistent with

this, a positive association between an unidentified species of *Flavobacterium* and semen quality was also observed (q=0.01, Table 3).

To focus analyses toward the most extreme phenotype of poor semen quality, a sub-analysis of controls compared with MFI was performed (Table 4). Non-parametric differential abundance analysis again identified a robust relationship between *Flavobacterium* and abnormal sperm morphology (q=0.01, Table 4). At species level, this was mapped to an unidentified species of *Flavobacterium* (q=0.01, Table 5). Similar to findings observed for all samples, sperm DNA fragmentation was inversely associated with relative abundance of *Cutibacterium* and positively associated with *Porphyromonas* and *Varibaculum* was also observed.

Discussion

We report the largest study to date investigating the seminal microbiota from patients suffering a range of adverse reproductive outcomes including unexplained infertility, male factor infertility and recurrent pregnancy loss. Moreover, we provide a detailed assessment of the relationship between semen microbial diversity, load and compositional structure with both molecular and classical seminal parameters. We identified 3 main clusters present in all study groups. Whilst overall bacterial composition was not association with aberrations in semen analysis, ROS and DNA fragmentation, men with unidentified *Flavobacterium* species were more likely to have abnormal semen quality or sperm morphology.

Several recent studies have indicated the existence of a semen microbiota, however these studies have been limited to small samples sizes and have failed to reach consensus on the compositional structure of the microbiota or its biological relevance, particularly in the context of sperm function and quality (13, 14, 16, 17, 18, 19, 34). By analysing the samples of 233 men with various reproductive disorders we offer a robust assessment of association between the microbiota and classical seminal parameters, but also key functional parameters including seminal ROS and sperm DNA damage. We incorporated stringent negative controls to permit removal of sequences likely originating from extraction kits and reagents known to contaminant low biomass samples such as semen (14, 16, 34). Molina et al report that 50%-70% of detected bacterial reads may be environmental contaminants in a sample from extracted testicular spermatozoa (35); with the addition of passage along the urethra it is likely that contamination of ejaculated semen would be much higher.

Mapping of genera level relative abundance data enabled semen samples to be categorised into 3 major clusters characterised by differing relative abundance of *Streptococcus*, *Prevotella*, *Lactobacillus* and *Gardnerella*. Unlike previous studies, we used an objective statistical approach (i.e. Silhouette methods) to determine the optimal number of microbial clusters supported by the data. These findings are largely consistent with earlier semen metataxonomic profiling studies reporting clusters enriched for *Streptococcus*,

Lactobacillus and *Prevotella* (13, 14, 16). Moreover, Baud *et al.*, reported increased bacterial richness in the *Prevotella*-enriched cluster, which we also observed (16). This may suggest that certain compositional characteristics of seminal microbiota are conserved across populations. However, similar modelling of species level data, failed to identify statistically robust clusters. This contrasts with other niches such as the vagina where reproducible clusters based on species level metataxonomic profiles have been demonstrated reflecting mutualistic relationships between specific species and the host, which have coevolved over long periods of time (36, 37). It is possible therefore that our findings indicate that microbiota detected in semen are likely the result of transient colonisation events. Consistent with this, several species known to be commensal to the penile skin including *Streptococcus*, *Corynebacterium* and *Staphylococcus*, or the female genital tract including *Gardnerella* and *Lactobacillus*, were observed in semen samples (38). This is in keeping with data suggesting microbiota transference during sexual intercourse (39). It remains possible that a proportion of bacteria detected in semen reflects contamination of the sample acquired during the collection procedure. Studies undertaking assessment of female partner microbiota profiles as well as temporal profiling of semen microbiota would improve understanding of potential dynamic restructuring of semen microbiota compositions. This has been done in part by Baud et al by studying the subfertile couple as a unit to establish if there is a ‘couple microbiota’(40) . They took samples from 65 couples with a range of pathologies including idiopathic infertility. From each woman they took vaginal swabs and follicular fluid samples. From each man they took a semen samples and penile swabs. They undertook extensive negative control series and stringent in silico elimination of possible contaminants. The found the male microbiota to be much more diverse than the female, with 90% of female samples being *Lactobacillus*-dominant. Intra-personal male samples i.e. semen and penile swabs from the same man bore more similarity to each other than inter-personal samples of the same sample type i.e. semen *or* penile swab comparisons between men (40). They identified that the male microbiota had very little impact of the microbiota of the female sexual partner(40) . Lack of information regarding the sexual activity of the enrolled couples limits this study somewhat.

Several previous studies have described semen microbiota composition to genera level and some have reported associations between specific genera and parameters of semen quality and function (13, 14, 16, 17, 18, 19, 34). However, in many cases these studies have failed to consider multiple comparisons testing, likely leading the reporting of spurious associations. We did not observe any significant associations between bacterial clusters, richness, diversity or load with traditional seminal parameters, sperm DNA fragmentation or semen ROS. This is in contrast with Veneruso *et al.*, who reported that in infertile patients, semen bacterial diversity and richness was decreased whereas Lundy *et al.*, reported that diversity was increased in infertile patients (17, 34). Further, Lundy et al., reported *Prevotella* abundance to be inversely associated with sperm concentration; this was not replicated in our study (17). There are several possible reasons accounting for the high heterogeneity in results including differences in methodology used to assess the microbial component of semen as well as differences in study design (41). For example, time of sexual abstinence prior to sample production as well as sample processing time often differs between studies, which has been shown to impact microbiological composition of semen (42).

The only association between bacterial taxa and semen parameters to withstand false detection rate testing for multiple comparisons detected in our study was between *Flavobacterium* and abnormal semen quality and sperm morphology (q=0.02). The Flavobacterium genus taxon we identified as significantly associated with abnormal semen quality and sperm morphology was present in 36.28% of the samples, with a mean relative abundance of 1.15% in those samples. This information and the mention of previous findings of Flavibacterium in contamination studies have been added to the discussion. *Flavobacterium* are gram-negative physiologically diverse aerobes, some of which are pathogenic (43). *Flavobacterium* was recently identified as a dominant genus in immature sperm cells retrieved from testicular biopsies of infertile men in a study by Molina et al (35). However, in contrast to these findings, a recent smaller study investigating semen collected from 14 sperm donors and 42 infertile idiopathic patients reported an association between *Flavobacterium* and increased sperm motility but a negative correlation with sperm DNA fragmentation (18). The genus Flavobacterium was defined in 1923 to encompass gram-negative, non-spore-forming rods,

of yellow pigment (44). The inclusiveness of this definition resulted in a collective of heterogenous species. By 1984 the genus had been restricted to those that were also non-motile and non-gliding (44). More recently, with an increase in genomic profiling, many species previously considered to be of genus Flavobacterium have been reclassified to genus Chryseobacterium, Cytophaga, and Weeksella (45). Increasing numbers of Flavobacterium species are being discovered such as *gondwanense, Collinsii, branchiarum, branchiicola, salegens* and *scophthalmum* (46) (47) (48). The allocation of Flavobacterium *aquatile* to this genus remains controversial due to its motility (49). Flavobacterium species are widely distributed in the environment including soil, fresh water and saltwater habitats (50) (51). There are many reports of pathogenic infections of Flavobacterium species in fish, however human infections are rare (48). A handful of case reports have described opportunistic infections to include pneumonia, urinary tract infection, peritonitis and meningitis (52) (53) (54) (55). Flavobacterium *lindanitolerans* and Flavobacterium *ceti* have been isolated as causative agents in some (56) (54). Case reports also describe Flavobacterium *odoratum* as a causative agent in urinary tract infection, most often in the immunocompromised or those with indwelling devices (57) (58) (59). However, this was one of many species previously of genus Flavobacterium reclassified, in this case to genus Myroides (60). Notably in our sample participants were asymptomatic of urinary tract infection.

Though not withstanding multiple correction, we did observe several other associations between specific bacterial taxa and semen parameters. For example, samples enriched with *Lactobacillus* had lower incidence of elevated seminal ROS, a relationship which could largely be accounted for by *Lactobacillus iners*, a common member of the cervicovaginal niche (61). Various studies have also found *Lactobacillus* enrichment in semen to associate with normal seminal parameters, especially morphology (14, 16). were *Lactobacillus*-predominant (14). However, an association between samples enriched with *Lactobacillus* and asthenospermia or oligoasthenospermia has also been described (19). We also observed an association between increased sperm DNA fragmentation and samples enriched with *Varibaculum*, which is consistent with previous reports of increased relative abundance of *Varibaculum* in semen infertile (34).

A limitation of this and other similar studies is that it was a single institutional study with limited ethnic diversity and potential geographical changes induced by environment or dietary habit. Gut microflora is known to display geographical variability; we cannot exclude that similar geographical variability exists for the seminal microbiota (62). The universal primers used during NGS may not be universal and may anneal variably to specific bacteria resulting in over-detection, under-detection, or indeed non-detection of some taxa (63) (64). A further limitation of this study, and others, is the lack of reciprocal genital tract microbiota testing of the female partners, or paired seminal and urinary samples from male participants. Additionally, we did not have other covariables such as smoking status with which to include in further analyses.

In summary, our study confirms that compositionally the semen microbiota can be broadly classified into three major groups based upon relative abundance of key bacterial genera. Despite different methodological approaches a number of studies, including our own, indicate a Prevotella-dominated or Lactobaccilus-dominated seminal microbiota, perhaps suggesting a stable microbiota at genera level. Our species level our data, however, failed to show similar clusters, perhaps instead suggesting transient colonisation. Longitudinal studies are required to ascertain the stability of the seminal microbiota. We provide evidence for an association between *Lactobacillus* abundance and normal seminal parameters. However, our results indicate that no specific semen bacterial composition can robustly differentiate between fertile and infertile men, although a small subset of bacteria may be associated with changes in seminal parameters. Our finding that an unidentified species of *Flavobacterium* impairs seminal parameters warrants further exploration and may offer the potential for targeted therapies. Larger, multi-centred studies as well as mechanistic investigations are required to establish causal links between the semen microbiota and male fertility.

Authors roles:

Mowla, Farahani, Tharakan, Jayasena and MacIntyre made substantial contribution to the study design, acquisition of data, analysis and interpretation of data and critical revision of the article for important intellectual content. Davies and Correia made substantial contribution to the analysis and interpretation of data and drafting the article. Lee, Kundu, Khanjani, Sindi and Khalifa made substantial contribution to the acquisition of data and critical revision of the article for important intellectual content. Rai, Regan, Henkel, Minhas Dhillo, Ben Nagi and Bennett made substantial contribution to the study design and critical revision of the article for important intellectual content. All authors approved the final version to be published and are in agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements: We would like to thank the patients and participants for their involvement in the study

Funding: The Section of Endocrinology and Investigative Medicine is funded by grants from the MRC, NIHR and is supported by the NIHR Biomedical Research Centre Funding Scheme and the NIHR/Imperial Clinical Research Facility. The views expressed are those of the author(s) and not necessarily those of the Tommy's, the NHS, the NIHR or the Department of Health. The following authors are also funded as follows: NIHR Research Professorship (WSD), NIHR Post-Doctoral Fellowship (CNJ). This project was supported by a research grant from Charm Foundation UK as well as funding by Tommy's National Centre for Miscarriage Research (grant P62774).

Conflict of interest: N/A

Data Availability Statement: The 16S rRNA metataxonomic dataset and the data analysis scripts are publicly available at the European Nucleotide Archive (Project accession PRJEB57401) and GitHub (repository link https://github.com/Gscorreia89/semen-microbiota-infertility), respectively).

References:

1. Levine H, Jorgensen N, Martino-Andrade A, Mendiola J, Weksler-Derri D, Mindlis I, et al. Temporal trends in sperm count: a systematic review and meta-regression analysis. Hum Reprod Update. 2017;23(6):646-59.
2. Eini F, Kutenaei MA, Zareei F, Dastjerdi ZS, Shirzeyli MH, Salehi E. Effect of bacterial infection on sperm quality and DNA fragmentation in subfertile men with Leukocytospermia. Bmc Mol Cell Biol. 2021;22(1).
3. Jung JH, Kim MH, Kim J, Baik SK, Koh SB, Park HJ, et al. Treatment of Leukocytospermia in Male Infertility: A Systematic Review. World J Mens Health. 2016;34(3):165-72.
4. Brunner RJ DJ, Sindhwani P. Review of guidelines for the evaluation and treatment of leukocytospermia in male infertility. World J Mens Health 2019;37:128-37.
5. Jung JH KM, Kim J, et al. . Treatment of leukocytospermia in male infertility: a systematic review. 2016. World J Mens Health 34:165–72.
6. Aitken RJ, Drevet JR, Moazamian A, Gharagozloo P. Male Infertility and Oxidative Stress: A Focus on the Underlying Mechanisms. Antioxidants-Basel. 2022;11(2).
7. Benchaib M, Lornage J, Mazoyer C, Lejeune H, Salle B, Guerin JF. Sperm deoxyribonucleic acid fragmentation as a prognostic indicator of assisted reproductive technology outcome. Fertil Steril. 2007;87(1):93-100.
8. Jayasena CN, Radia UK, Figueiredo M, Revill LF, Dimakopoulou A, Osagie M, et al. Reduced Testicular Steroidogenesis and Increased Semen Oxidative Stress in Male Partners as Novel Markers of Recurrent Miscarriage. Clin Chem. 2019;65(1):161-9.
9. Aitken RJ, Bakos HW. Should we be measuring DNA damage in human spermatozoa? New light on an old question. Hum Reprod. 2021;36(5):1175-85.
10. De Iuliis GN, Thomson LK, Mitchell LA, Finnie JM, Koppers AJ, Hedges A, et al. DNA Damage in Human Spermatozoa Is Highly Correlated with the Efficiency of Chromatin Remodeling and the Formation of 8-Hydroxy-2 '-Deoxyguanosine, a Marker of Oxidative Stress. Biol Reprod. 2009;81(3):517-24.
11. Vorilhon S, Brugnon F, Kocer A, Dollet S, Bourgne C, Berger M, et al. Accuracy of human sperm DNA oxidation quantification and threshold determination using an 8-OHdG immuno-detection assay. Hum Reprod. 2018;33(4):553-62.
12. Agarwal A, Sharma RK, Nallella KP, Thomas AJ, Alvarez JG, Sikka SC. Reactive oxygen species as an independent marker of male factor infertility. Fertil Steril. 2006;86(4):878-85.
13. Hou DS, Zhou X, Zhong X, Settles ML, Herring J, Wang L, et al. Microbiota of the seminal fluid from healthy and infertile men. Fertil Steril. 2013;100(5):1261-+.
14. 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(10).
15. Monteiro C, Marques PI, Cavadas B, Damiao I, Almeida V, Barros N, et al. Characterization of microbiota in male infertility cases uncovers differences in seminal hyperviscosity and oligoasthenoteratozoospermia possibly correlated with increased prevalence of infectious bacteria. Am J Reprod Immunol. 2018;79(6).
16. Baud D PC, Vulliemoz N, Castella V, Marsland BJ, Stojanov M. Sperm Microbiota and Its Impact on Semen Parameters. Frontiers in Microbiology. 2019;10:234.
17. Lundy SD, Sangwan N, Parekh NV, Selvam MKP, Gupta S, McCaffrey P, et al. Functional and Taxonomic Dysbiosis of the Gut, Urine, and Semen Microbiomes in Male Infertility. Eur Urol. 2021;79(6):826-36.
18. Garcia-Segura S DRJ, Closa L, Garcia-Martínez I, Hobeich C, Castel AB, Vidal F, Benet J, Ribas-Maynou J, Oliver-Bonet M. Seminal Microbiota of Idiopathic Infertile Patients and Its Relationship With Sperm DNA Integrity. Front Cell Dev Biol. 2022;28(10):937157.
19. Yang H ZJ, Xue Z, Zhao C, Lei L, Wen Y, Dong Y, Yang J, Zhang L. Potential Pathogenic Bacteria in Seminal Microbiota of Patients with Different Types of Dysspermatism. Sci Rep. 2020;23(10):6876.
20. Altmäe S FJ, Mändar R. The seminal microbiome in health and disease. Nat Rev Urol 2019;16.
21. World Health Organization. WHO laboratory manual for the examination and processing of human semen. 5th ed. Geneva: WHO Press.

22. Scheme UNRS. UK NEQAS Reproductive Science Scheme Department of Reproductive Medicine Old Saint Mary's Hospital
Oxford Road
Manchester, M13 9WL
United Kingdom2023 [Scheme Organiser – Ms. Annette Lloyd]. Available from: https://mft.nhs.uk/saint-marys/services/gynaecology/reproductive-medicine/ukneqasreproductive-science/.
23. Vessey W, Perez-Miranda A, Macfarquhar R, Agarwal A, Homa S. Reactive oxygen species in human semen: validation and qualification of a chemiluminescence assay. Fertil Steril. 2014;102(6):1576-U389.
24. Sergerie M LG, Bujan L et al. Sperm DNA fragmentation: threshold value in male fertility. Hum Reprod. 2005;20(12).
25. Sharma R IC, Agarwal A, Henkel R. TUNEL assay-Standardized method for testing sperm DNA fragmentation. Andrologia. 2020;53(2):e13738.
26. Hughes CM LS, McKelvey-Martin VJ, Thompson W. A comparison of baseline and induced DNA damage in human spermatozoa from fertile and infertile men, using a modified comet assay. MHR: Basic science of reproductive medicine. 1996;2(8):613-9.
27. Liu CM, Aziz M, Kachur S, Hsueh P-R, Huang Y-T, Keim P, et al. BactQuant: An enhanced broad-coverage bacterial quantitative real-time PCR assay. BMC Microbiology. 2012;12(1):56.
28. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-7.
29. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.
30. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
31. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6(1):226.
32. Nunez-Calonge R CP, Redondo C. Urealyticum Reduces Motility and Induces Membrane Alterations in Human Spermatozoa. Hum Reprod 1998;13:2756–61.
33. Satta A SA, Garozzo A. Experimental Chlamydia Trachomatis Infection Causes Apoptosis in Human Sperm. Hum. Reprod 2006;21:134–7.
34. Veneruso I CF, Alviggi C, Pastore L, Tomaiuolo R, D'Argenio V. Metagenomics Reveals Specific Microbial Features in Males with Semen Alterations. Genes (Basel). 2023;14(6):1228.
35. Molina NM, Plaza-Diaz J, Vilchez-Vargas R, Sola-Leyva A, Vargas E, Mendoza-Tesarik R, et al. Assessing the testicular sperm microbiome: a low-biomass site with abundant contamination. Reprod Biomed Online. 2021;43(3):523-31.
36. Pruski P, Correia GDS, Lewis HV, Capuccini K, Inglese P, Chan D, et al. Direct on-swab metabolic profiling of vaginal microbiome host interactions during pregnancy and preterm birth. Nature Communications. 2021;12(1):5967.
37. France MT, Ma B, Gajer P, Brown S, Humphrys MS, Holm JB, et al. VALENCIA: a nearest centroid classification method for vaginal microbial communities based on composition. Microbiome. 2020;8(1):166.
38. Byrd AL, Belkaid Y, Segre JA. The human skin microbiome. Nature Reviews Microbiology. 2018;16(3):143-55.
39. Ma Z. Microbiome Transmission During Sexual Intercourse Appears Stochastic and Supports the Red Queen Hypothesis. Frontiers in Microbiology. 2022;12.
40. Baud D PA, Vidal JM et al Interaction of genital microbiota in infertile couples. bioRxiv 20230614544778; doi: https://doiorg/101101/20230614544778. 2023

.
41. Farahani L, Tharakan T, Yap T, Ramsay JW, Jayasena CN, Minhas S. The semen microbiome and its impact on sperm function and male fertility: A systematic review and meta-analysis. Andrology. 2021;9(1):115-44.
42. Yao TH, X.; Guan, T.; Wang, Z.; Zhang, S.; Liu, C.; Liu, C.; Chen, L. Effect of indoor environmental exposure on seminal microbiota and its application in body fluid identification. Forensic Sci Int. 2020;314:110417.

43. Waskiewicz A IL. Flavobacterium spp. – Characteristics, Occurrence, and Toxicity. In: Batt CA TM, editor. Encyclopedia of Food Microbiology 2nd ed. Oxford: Oxford: Academic Press; 2014. p. 938-42.
44. Holmes B OR, McMeekin TA Genus Flavobacterium Bergey, Harrison, Breed, Hammer and Huntoon 1923. In: Krieg N. R. HJG, editor. Bergey's manual of systematic bacteriology. Baltimore: The Williams and Wilkins Co; 1984.
45. Bernardet JF SP, Vancanneyt M et al. Cutting a Gordian Knot: Emended Classification and Description of the Genus Flavobacterium, Emended Description of the Family Flavobacteriaceae, and Proposal of Flavobacterium hydatis nom. nov. (Basonym, Cytophaga aquatilis Strohl and Tait 1978). International Journal of Systematic and Evolutionary Microbiology 1996;International Journal of Systematic and Evolutionary Microbiology 128 - 48.
46. Dobson SJ CR, McMeekin TA, Franzmann PD 1993; Direct sequencing of the polymerase chain reaction-amplified 16S rRNA gene of Flavobaderium gondwanense sp. nov. and Flavobacterium salegens sp. nov., two new species from a hypersaline antarctic lake. Int. J. Syst. Bacteriol 43:77–83. Direct sequencing of the polymerase chain reaction-amplified 16S rRNA gene of Flavobaderium gondwanense sp. nov. and Flavobacterium salegens sp. nov., two new species from a hypersaline antarctic lake. Int J Syst Bacteriol 1993;43:77-83.
47. Mudarris M AB, Segers P Flavobacterium scophthalmum sp. nov., a pathogen of turbot (Scophthalmus maximus L.). Int J Syst Bacteriol 1994;44:447–53.
48. Zamora L VA, Sánchez-Porro C et sl. Characterization of flavobacteria possibly associated with fish and fish farm environment. Description of three novel Flavobacterium species: Flavobacterium collinsii sp. nov., Flavobacterium branchiarum sp. nov., and Flavobacterium branchiicola sp. nov. Aquaculture. 2013;416–417:346-53.
49. Sheu SY LY, Chen WM. Flavobacterium squillarum sp. nov., isolated from a freshwater shrimp culture pond, and emended descriptions of Flavobacterium haoranii, Flavobacterium cauense, Flavobacterium terrae and Flavobacterium aquatile. Int J Syst Evol Microbiol 2013;63:2239–47.
50. Loch TP FM. Emerging flavobacterial infections in fish: a review. J Adv Res 2015;6:283–300.
51. Vela AI FA, Sánchez-Porro C Flavobacterium ceti sp. nov., isolated from beaked whales (Ziphius cavirostris). Int J Syst Evol Microbiol 2007;57:2604–8.
52. Sung JY KT, Shin S et al. Flavobacterium ceti from blood samples of a Korean patient with alcoholic liver cirrhosis. Ann Lab Med 2015;35:384–6.
53. Tian GZ PD, Zhao HY et al. A Flavobacterium lindanitolerans strain isolated from the ascites sample of a Chinese patient with EV71 virus infection. Biomed Environ Sci 2011;24:694–6.
54. Park SK RN. A case of Flavobacterium ceti meningitis. Ann Lab Med 2016;36:614–6.
55. Mosayebi Z MA, Soori T. Flavobacterium sepsis outbreak due to contaminated distilled water in a neonatal intensive care unit. J Hosp Infect 2011;78:214–5.
56. Zurbuchen R dRM, Galimanis A et al. First Case of Meningoencephalitis and Bacteraemia with Flavobacterium Lindanitolerans. Eur J Case Rep Intern Med 2023;13(10):003980.
57. Benedetti P RM, Pavan G Septic shock, pneumonia, and soft tissue infection due to Myroides odoratimimus: report of a case and review of Myroides infections. Infection 2011;39(2):161-5.
58. Chauhan K CP, Singh R, et al. Myroides causing catheter associated urinary tract infection in diabetic patients: an emerging multidrug resistant "Superbug". J Clin Diagn 2020;14(11):1-4.
59. Verma Y HS, Shah K Urinary tract infection caused by Myroides spp. in diabetic patients: To be or not to be. Indian J Case Reports 2018;4(1):1-3.
60. Vancanneyt M SP, Torck U et al. Reclassification of Flavobacterium odoratum(Stutzer 1929) strains to a new genus, Myroides, as Myroides odoratus comb. nov. and Myroides odoratimimus sp. nov. Int J Syst Bacteriol 1996;46(9):26–32.
61. MacIntyre DA, Sykes L, Bennett PR. The human female urogenital microbiome: complexity in normality. Emerging Topics in Life Sciences. 2017;1(4):363-72.
62. Suzuki TA, Worobey M. Geographical variation of human gut microbial composition. Biol Letters. 2014;10(2).
63. Forney LJ, Zhou X, Brown CJ. Molecular microbial ecology: land of the one-eyed king. Curr Opin Microbiol. 2004;7(3):210-20.
64. von Wintzingerode F, Gobel UB, Stackebrandt E. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. Fems Microbiol Rev. 1997;21(3):213-29.

Figures

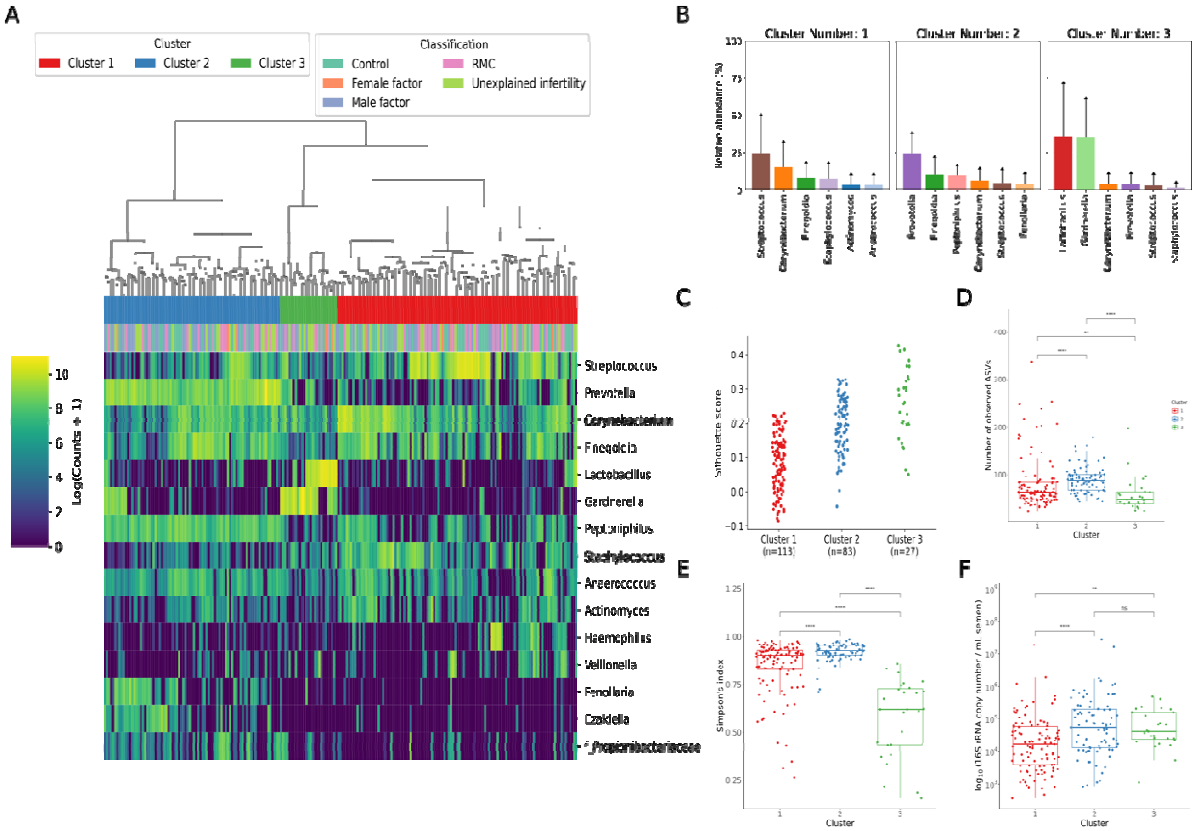

Figure 1. Characterisation of semen microbiota composition at genera level. A) Heatmap of Log10 transformed read counts of top 10 most abundant genera identified in semen samples. Samples clustered into three major microbiota groups based mainly on dominance by *Streptococcus* (Cluster 1), *Prevotella* (Cluster 2), or *Lactobacillus* and *Gardnerella* (Cluster 3). (n=223, Ward's linkage). **B)** Relative abundance of the top 6 most abundant genera within each cluster. **C)** Silhouette scores of individual samples within each cluster. **D)** Species richness ($p<0.0001$; Kruskal-Wallis test) and **E**) alpha diversity ($p<0.0001$; Kruskal-Wallis test) significantly differed across clusters. **F**) Assessment of bacterial load using qPCR showed Clusters 2 and 3 have significantly higher bacterial loads compared to Cluster 1 Dunn's multiple comparison test was used as a post-hoc test for between group comparisons (*$p<0.05$, ****$p<0.0001$).

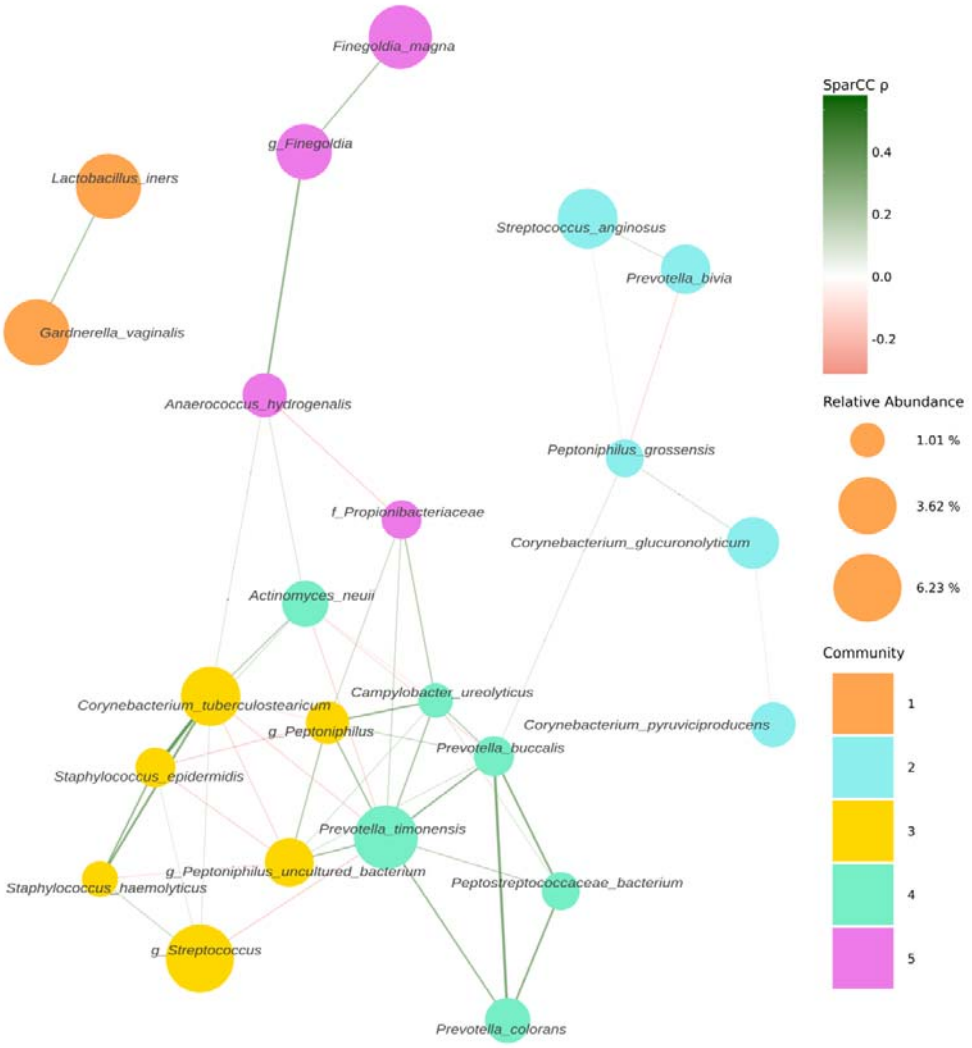

Figure 2. Co-occurrence network estimated with SparCC from 16S sequencing counts at species level. Network representing co-occurrence patterns (edges), between various taxonomic units, assigned at species level (nodes). Edges are colored by their estimated SparCC correlation coefficient (ρ). Edges with a SparCC bootstrapped *p-value* < 0.05, ρ < 0.25, and singleton nodes are not shown. Node color represents network community membership. Node sizes are proportional to the mean relative abundance of their respective taxon.

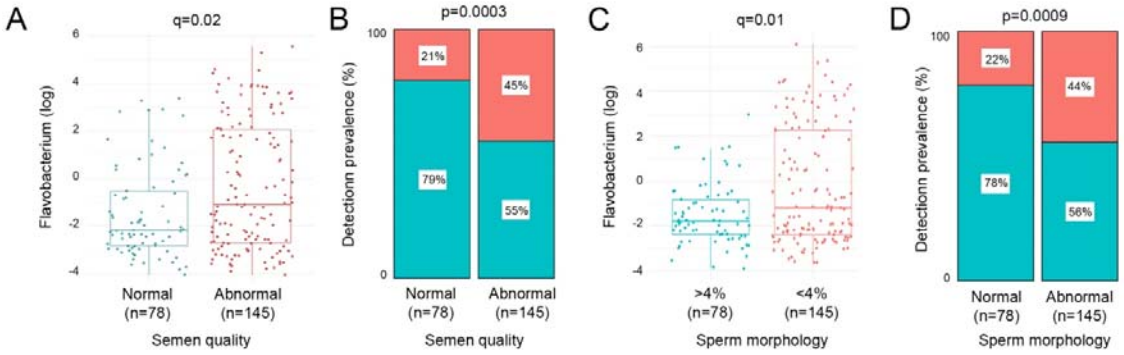

Figure 3. Relative abundance and prevalence matrices of Flavobacterium in relation to semen quality and morphology. **A)** Relative abundance of Flavobacterium was significantly higher in samples with abnormal semen (p=0.0002, q=0.02). **B)** Detection of flavobacterium was significantly more prevalent in abnormal semen quality samples (p=0.0003). **C)** Flavobacterium relative abundance was significantly higher in samples with <4% morphologically normal forms (p=0.0002, q=0.01). **D)** Flavobacterium was also significantly more prevalent in samples with low percentage of morphologically normal sperm (p=0.0009).

Tables:

Table I: Patient demographics and notable parameters of seminal quality and function for controls and study subjects. Fisher's exact tests for all except age. Chi-square test for age. (n=223).

Factor	Categories	Controls	Study cases	*p-value*
DNA fragmentation index	Low	45/114 (40%)	69/114 (60%)	0.0002***
	High	12/ 82 (15%)	70/82 (85%)	
ROS	<3.77 RLU/s	53/143 (37%)	90/143 (63%)	0.02*
	>3.77 RLU/s	5/33 (15%)	28/33 (85%)	
Semen volume	Optimal	55/208 (26%)	153/208 (84%)	0.03*
	Sub-optimal	8/15 (53%)	7/15 (47%)	
Age	<34	11/49 (22%)	38/49 (88%)	0.04*
	34-41	31/124 (25%)	93/124 (85%)	
	>41	21/50 (42%)	29/50 (58%)	
Ethnicity	Caucasian	39/156 (25%)	117/156 (75%)	0.10
	Non-Caucasian	24/67 (36%)	43/67 (64%)	
Concentration	>15 M/mL	58/182 (32%)	124/182 (68%)	0.01*
	<15 M/mL	5/41 (21%)	36/41 (79%)	
Progressive motility	>32%	60/207 (29%)	147/207 (71%)	0.56
	<32%	3/16 (19%)	13/16 (81%)	
Sperm morphology	>4%	22/74 (30%)	52/74 (70%)	0.87
	<4%	41/144 (28%)	103/144 (72%)	
Semen quality	Optimal	24/78 (31%)	54/78 (69%)	0.53
	Sub-optimal	39/145 (27%)	106/145 (73%)	

Table II: Differential abundance analysis for bacterial genera with seminal quality and functional parameters. Positive t-values indicate a positive relationship, and a negative t-value describes a negative relationship between relative abundance of taxa and seminal quality and function parameters. Significant relationships are indicated using p-values. q-values represent Benjamini-Hochberg false discovery rate corrected p-values for multiple comparisons.

Sperm quality and function parameters	**Genera**	**Welch's *t-statistic***	***p-value***	***q-value***
Sperm DNA fragmentation	*Finegoldia*	-2.36	0.01*	0.27
	Cutibacterium	-2.20	0.02*	0.27
	Porphyromonas	2.16	0.03*	0.27
	Varibaculum	2.11	0.03*	0.27
ROS	*Lactobacillus*	2.18	0.02*	0.66
	Corynebacterium	-2.04	0.04*	0.66
Semen quality	*Flavobacterium*	3.39	0.0008***	0.02*
	Prevotella	2.26	0.02*	0.38
Sperm concentration	*Porphyromonas*	-2.08	0.03*	0.61
Sperm morphology	*Flavobacterium*	3.64	0.0003***	0.01*
	Prevotella	2.03	0.04*	0.67
Semen volume	*Corynebacterium*	2.27	0.02*	0.32
	Actinotigum	-2.20	0.02*	0.32
	Varibaculum	-2.16	0.03*	0.32

Table III: Differential abundance analysis for bacterial species with seminal quality and functional parameters. Positive t-values indicate a positive relationship and a negative t-value describes a negative relationship between relative abundance of taxa and seminal quality and function parameters. Significant relationships are indicated using p-values. q-values represent Benjamini-Hochberg false discovery rate corrected p-values for multiple comparisons.

Clinical factor	**Species**	**Welch's *t-statistic***	***p-value***	***q-value***
Sperm DNA fragmentation	*Peptostreptococcaceae bacterium*	2.18	0.03*	0.91
ROS	*Lactobacillus iners*	2.24	0.02*	0.94
	Unidentified Anaerococcus	-2.03	0.04*	0.94
Semen quality	Unidentified Flavobacterium	3.76	0.0002***	0.01*
	Corynebacterium tuberculostearicum	-2.06	0.04*	0.82
Semen volume	*Corynebacterium tuberculostearicum*	2.64	0.008	0.24
	Unidentified Varibaculum	-2.48	0.01	0.24
	Staphylococcus epidermidis	2.35	0.01	0.24
	Unidentified Peptoniphilus	-2.32	0.02	0.24
	Dialister propionicifaciens	-2.24	0.02	0.24
	Prevotella colorans	-2.14	0.03	0.26
Cohorts	*Staphylococcus haemolyticus*	0.04	0.02	0.97

Table IV: Differential abundance analysis for specific taxa at genera level for controls and cases with male factor infertility. Positive t-values indicate a relationship, and a negative t-value describes a negative relationship between relative abundance of taxa and seminal quality and function parameters. Significant relationships are indicated using p-values. q-values represent Benjamini-Hochberg false discovery rate corrected p-values for multiple comparisons.

Clinical factor	**Genera**	**Welch's *t-statistic***	***p-value***	***q-value***
Sperm DNA fragmentation	*Cutibacterium*	-2.56	0.01*	0.31
	Porphyromonas	2.34	0.02*	0.31
	Varibaculum	1.96	0.051	0.53
ROS	*Finegoldia*	-1.99	0.04*	0.77
Sperm concentration	*Finegoldia*	2.04	0.04*	0.71
Sperm morphology	*Flavobacterium*	3.64	0.0003***	0.01*
	Prevotella	2.03	0.04*	0.67
Semen volume	*Facklamia*	2.99	0.003**	0.10
	Actinotignum	-2.20	0.02*	0.36
	Dialister	-1.99	0.04*	0.36

Table V: Differential abundance analysis for specific taxa at species for controls and male factor infertility. Positive t-values indicate a positive relationship and a negative t-value describes a negative relationship between relative abundance of taxa and seminal quality and function parameters. Significant relationships are indicated using p-values. q-values represent Benjamini-Hochberg false discovery rate corrected p-values for multiple comparisons.

Clinical factor	**Species**	**Welch's *t-statistic***	***p-value***	***q-value***
Sperm DNA fragmentation	*Staphylococcus hominis*	-2.32	0.02*	0.68
ROS	Unidentified Flavobacterium	2.42	0.01	0.54
	Unidentified Anaerococcus	-2.12	0.03	0.54
	Schaalia radingae	-2.12	0.03*	0.54
	Haemophilus parainfluenza	2.02	0.04*	0.54
Semen quality	Unidentified Flavobacterium	2.36	0.01*	0.91
Semen volume	*Dialister micraerophilus*	-2.66	0.008**	0.41
	Corynebacterium tuberculostearicum	2.27	0.02*	0.44
	Staphylococcus epidermidis	2.22	0.02*	0.44
	Actinotignum schaalii	-2.00	0.04*	0.45
Cohorts	*Staphylococcus haemolyticus*	0.04	0.01*	0.68