

## SEMEN MICROBIOTA: CLUSTER ANALYSIS OF REAL-TIME PCR DATA

Voroshilina ES<sup>1,2</sup>✉, Zornikov DL<sup>1</sup>, Ivanov AV<sup>3</sup>, Pochernikov DG<sup>5</sup>, Panacheva EA<sup>1,2</sup><sup>1</sup> Ural State Medical University, Yekaterinburg, Russia<sup>2</sup> Medical Center "Garmonia", Yekaterinburg, Russia<sup>3</sup> Yeltsin Ural Federal University, Yekaterinburg, Russia<sup>4</sup> Krasovskii Institute of Mathematics and Mechanics, Yekaterinburg, Russia<sup>5</sup> Ivanovo State Medical Academy, Ivanovo, Russia

To this day semen microbiota is still poorly understood, and clinical significance of detecting specific microorganism groups has not been clearly determined. The aim of this work was to conduct cluster analysis of semen microbiota detected using real-time PCR. 634 semen samples of reproductive age men were analyzed using the Androflor kit. Microbial DNA in the quantity of no less than  $10^3$  GE/ml was detected in 460 samples (72.5%). From 1 to 14 microorganism groups were detected in 350 samples (55.2%) in the quantities that exceeded the threshold values (the detection rate of specific groups: 3.3–21.0%). In these 350 samples 4 stable microbiota clusters were determined. Each of the clusters was characterized by the prevalence of a specific microorganism group: obligate anaerobes (cluster 1;  $n = 172$ ; detection rate — 49.1%), *Lactobacillus spp.* (cluster 2;  $n = 78$ ; detection rate — 22.3%), gram-positive facultative anaerobes (cluster 3;  $n = 62$ ; detection rate — 17.7%), *Enterobacteriaceae / Enterococcus* (cluster 4;  $n = 62$ ; detection rate — 10.9%). Cluster 1 was less stable and was characterized by the larger species diversity compared to other clusters.

**Keywords:** semen microbiota, real-time PCR, cluster analysis, semen analysis

**Acknowledgments:** the authors would like to thank VN Khayutin, director of "Garmonia" Medical Center, for allowing them to conduct the study in the clinic's laboratory department.

**Author contribution:** Voroshilina ES — organization of the study, data analysis, article authoring; Zornikov DL — data analysis, article authoring; Ivanov AV — statistical processing, data analysis, article authoring; Pochernikov DG — patients' clinical profile, clinical data collection, data analysis, article authoring; Panacheva EA — literature review, data analysis, conducting PCR tests, article authoring.

**Compliance with ethical standards:** the study was approved by the Ethics Committee of Ural State Medical University, Federal State Budget Educational Institution of Higher Education under the Ministry of Health of the Russian Federation (Protocol № 7 of September 20, 2019). All patients signed the informed written consent to participation in the study.

✉ **Correspondence should be addressed:** Ekaterina S. Voroshilina  
Furmanova, 30, Yekaterinburg, 620142; voroshilina@gmail.com

**Received:** 21.09.2020 **Accepted:** 08.10.2020 **Published online:** 27.10.2020

**DOI:** 10.24075/brsmu.2020.064

## МИКРОБИОТА ЭЯКУЛЯТА: КЛАСТЕРНЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ, ПОЛУЧЕННЫХ ПРИ ИССЛЕДОВАНИИ МЕТОДОМ ПЦР-РВ

Е. С. Ворошилина<sup>1,2</sup>✉, Д. Л. Зорников<sup>1</sup>, А. В. Иванов<sup>3,4</sup>, Д. Г. Почерников<sup>5</sup>, Е. А. Паначева<sup>1,2</sup><sup>1</sup> Уральский государственный медицинский университет, Екатеринбург, Россия<sup>2</sup> Медицинский центр «Гармония», Екатеринбург, Россия<sup>3</sup> Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, Екатеринбург, Россия<sup>4</sup> Институт математики и механики имени Н. Н. Красовского Уральского отделения Российской академии наук, Екатеринбург, Россия<sup>5</sup> Ивановская государственная медицинская академия, Иваново, Россия

Микробиота эякулята до сих пор недостаточно изучена, а клиническая значимость выявления отдельных групп микроорганизмов окончательно не установлена. Целью работы было провести кластерный анализ микробиоты эякулята, выявленной методом ПЦР-РВ. С помощью теста «Андрофлор» («ДНК-Технология») исследовали 634 образца эякулята мужчин репродуктивного возраста. Микробную ДНК в количестве не менее  $10^3$  ГЭ/мл обнаружили в 460 (72,5%) образцов. В 350 (55,2%) пробах выявили от 1 до 14 групп микроорганизмов в надпороговых значениях (частота выявления отдельных групп составила 3,3–21,0%). Среди 350 образцов выделили четыре устойчивых кластера микробиоты, в каждом из которых преобладала определенная группа микроорганизмов: облигатных анаэробов (кластер 1,  $n = 172$ ; частота выявления — 49,1%), *Lactobacillus spp.* (кластер 2,  $n = 78$ ; частота выявления — 22,3%) грамположительных факультативных анаэробов (кластер 3,  $n = 62$ ; частота выявления — 17,7%), *Enterobacteriaceae / Enterococcus* (кластер 4,  $n = 38$ ; частота выявления — 10,9%). Кластер 1 характеризовался меньшей устойчивостью и большим видовым разнообразием в сравнении с другими кластерами.

**Ключевые слова:** микробиота эякулята, ПЦР-РВ, кластерный анализ, спермограмма

**Благодарности:** авторы благодарят директора медицинского центра «Гармония» (г. Екатеринбург) В. Н. Хаютина за возможность выполнения исследования на базе центра.

**Вклад авторов:** Е. С. Ворошилина — организация исследования, анализ данных, написание статьи; Д. Л. Зорников — анализ данных, написание статьи; А. В. Иванов — статистическая обработка, анализ данных, написание статьи; Д. Г. Почерников — клиническая характеристика пациентов, сбор клинических данных, анализ данных, написание статьи; Е. А. Паначева — анализ литературы, проведение ПЦР, анализ данных, написание статьи.

**Соблюдение этических стандартов:** исследование одобрено этическим комитетом Уральского государственного медицинского университета (протокол № 7 от 20 сентября 2019 г.). Все участники исследования подписали добровольное информированное согласие на проведение исследования.

✉ **Для корреспонденции:** Екатерина Сергеевна Ворошилина  
ул. Фурманова, д. 30, г. Екатеринбург, 620142; voroshilina@gmail.com

**Статья получена:** 21.09.2020 **Статья принята к печати:** 08.10.2020 **Опубликована онлайн:** 27.10.2020

**DOI:** 10.24075/vrgmu.2020.064

Male genital tract microbiota and semen microbiota in particular are still poorly understood compared to the microbiota of other human body biotopes [1]. For a long time, semen in healthy men was considered to be sterile, and any microorganisms (MO) detected there were associated with pathologies. Nevertheless, recent studies indicate that microbiota can be present in the semen of healthy or asymptomatic men with normal semen parameters [1–7]. It has been shown that polymicrobial communities of various bacteria genera and even phyla constitute semen microbiota [1, 2, 5, 7]. Some authors even cautiously conclude that certain MO groups could be associated with norm and pathology [1, 2, 5]. There are also researchers who believe that it is the presence of certain microbial associations, not species, that is associated with genital tract inflammatory diseases [4].

These results became possible due to the implementation of molecular-based techniques since many of the microbes detected in semen are difficult to culture or non-culturable (including obligate anaerobic bacteria which are rarely found in a routine culture-based test) [4, 7, 8]. However, clinical significance of the detection of these MO in semen samples has not been clearly established.

Most of the research dedicated to analyzing semen microbiota are based on 16S rRNA gene specific Next generation sequencing [1–5, 7]. While it is highly informative, this approach has a number of disadvantages such as: complicated sample preparation, difficult sample intake control, complicated result interpretation, long analysis process, high cost of equipment and reagents. These disadvantages make using NGS-sequencing in routine medical practice virtually impossible. Quantitative real-time PCR (qPCR) is far more suitable for this. In several previous studies, the potential of the Androflor commercial kit (qPCR kit for the detection of 24 MO groups) for semen microbiota analysis was shown [9–11]. Among other things, Androflor kit is more informative compared to culture-based tests [10]. While analyzing semen microbiota using qPCR has many benefits compared to other microbiological techniques, practical interpretation of the results remains difficult, which prevents this method from becoming part of the routine practice.

Semen culture colony count of  $10^3$  CFU/1 ml or higher is considered to be the above the threshold value for detecting opportunistic microbiota in culture-based testing [12]. The high sensitivity of molecular-based techniques and their capability to detect non-culturable and non-viable MO makes it difficult to use threshold values similar to those used in culture-based techniques when interpreting qPCR results. It is necessary to establish whether the presence of non-culturable MOs in quantities exceeding the threshold value is typical for normal and pathological conditions. We also need to determine the persistent types of microbial groups associated with infertility in men when identifying certain MO groups.

To answer these questions, semen analysis results (clinical and molecular-based), both from patients with infertility and healthy males, need to be comprehensively studied. Aim of the study: to conduct cluster analysis of semen microbiota detected by means of real-time PCR (Androflor kit)

## METHODS

### Patient groups

From January 2019 to March 2020, semen samples from 634 men were examined (mean age  $34 \pm 6.7$  years). During this period, the patients came to the “Garmonia” Medical Center

(Yekaterinburg,  $n = 429$ ) and to the urological clinic of the Ivanovo State Medical Academy of the Ministry of Health of Russian Federation (Ivanovo,  $n = 205$ ) either seeking preconception care or for infertility treatment. All patients gave their consent to participate in the study.

Inclusion criteria: reproductive-age men; infertility or undergoing preconception care; all examined patients during the last four weeks did not receive medications that could affect the semen microbiota, such as hormonal, antibacterial drugs; consumption of substances with alcohol content over 30 ml in terms of pure ethanol was excluded.

Exclusion criteria: hypogonadotropic and hypergonadotropic hypogonadism, type 1 and 2 diabetes, hypo- and hyperthyroidism; sexually transmitted infections (*Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Mycoplasma genitalium*, *Trichomonas vaginalis*); clinical manifestations of prostatitis such as pain and dysuria; karyotype abnormalities, mutations in the CFTR gene, microdeletions in the AZF locus of the Y chromosome.

### Semen sampling

Preparation for semen sampling: sexual abstinence for a period of 2–5 days. Prior to semen collection, patients urinated. Semen was collected through masturbation into a sterile container. Patients were instructed to avoid contact with the walls and the lid of the container.

### DNA extraction

PREP-NA-PLUS kit (DNA-Technology; Russia) was used for DNA-extraction. Semen samples were prepared using the following technique: 1.0 ml of semen was put into an Eppendorf tube with 1.0 ml of transport medium (“Transport media with mucolytic agent”, InterLabService Ltd.; Russia) which was then shaken in the vortex until the substances mixed completely. The tube was centrifuged at 13,000 rpm for 10 minutes (Mini-Spin centrifuge, Eppendorf; Germany). After removing the supernatant, 50  $\mu$ l of the precipitate was used for extraction of the DNA.

### Semen microbiota evaluation

The study was conducted using the Androflor reagent kit (DNA-Technology; Russia) and the DTprime detection thermal cycler (DNA-Technology; Russia) following the manufacturer's instructions. Once the amplification is over, the special software (DNA-Technology; Russia) automatically calculates the quantities (expressed in genome equivalents per 1 ml (GE/ml)) of the total bacterial load (TBL), lactobacilli and each of the detected opportunistic microorganisms (OM) in a given sample.

The Androflor kit allows detecting the following MO groups: Gram-positive facultative anaerobes (*Streptococcus spp.*, *Staphylococcus spp.*, *Corynebacterium spp.*); Gram-negative facultative anaerobes (*Haemophilus spp.*, *Pseudomonas aeruginosa* / *Ralstonia spp.* / *Burkholderia spp.*); *Enterobacteriaceae* / *Enterococcus spp.* group; obligate anaerobes (*Gardnerella vaginalis*, *Eubacterium spp.*, *Sneathia spp.* / *Leptotrichia spp.* / *Fusobacterium spp.*, *Megasphaera spp.* / *Veillonella spp.* / *Dialister spp.*, *Bacteroides spp.* / *Porphyromonas spp.* / *Prevotella spp.*, *Anaerococcus spp.*, *Peptostreptococcus spp.*, *Atopobium cluster*), mycoplasmas (*Mycoplasma hominis*, *Ureaplasma urealyticum*, *Ureaplasma parvum*), transient microbiota (*Lactobacillus spp.*), yeast-like fungi (*Candida spp.*).

**Table 1.** Detection rate for individual MO groups present in quantities exceeding the threshold value ( $n = 634$ )\*

| MO Group   | $n$ | %    |
|--|-----|------|
| <i>Corynebacterium spp.</i>  | 133 | 21.0 |
| <i>Bacteroides spp. / Porphyromonas spp. / Prevotella spp.</i>     | 131 | 20.7 |
| <i>Lactobacillus spp.</i>  | 125 | 19.7 |
| <i>Eubacterium spp.</i>  | 108 | 17.0 |
| <i>Peptostreptococcus spp. / Parvimonas spp.</i>                   | 107 | 16.9 |
| <i>Megasphaera spp. / Veillonella spp. / Dialister spp.</i>        | 94  | 14.8 |
| <i>Streptococcus spp.</i>  | 81  | 12.8 |
| <i>Enterobacteriaceae spp. / Enterococcus spp.</i>                 | 74  | 11.7 |
| <i>Anaerococcus spp.</i>   | 68  | 10.7 |
| <i>Gardnerella vaginalis</i>                                       | 67  | 10.6 |
| <i>Ureaplasma parvum</i>   | 63  | 9.9  |
| <i>Atopobium cluster</i>   | 57  | 9.0  |
| <i>Staphylococcus spp.</i>   | 52  | 8.2  |
| <i>Sneathia spp. / Leptotrichia spp. / Fusobacterium spp.</i>      | 41  | 6.5  |
| <i>Haemophilus spp.</i>  | 36  | 5.7  |
| <i>Mycoplasma hominis</i>  | 25  | 3.9  |
| <i>Ureaplasma urealyticum</i>                                      | 24  | 3.8  |
| <i>Pseudomonas aeruginosa / Ralstonia spp. / Burkholderia spp.</i> | 21  | 3.3  |

**Note:** \* — for *Ureaplasma urealyticum*, *Ureaplasma parvum*, *Mycoplasma hominis* threshold is  $> 0$ , for the other MO groups threshold is  $\geq 10^3$  GE/ml.

Sterile deionized water was used as the negative control sample. Positive signals were detected in the negative control sample for some MO groups no earlier than in the 35<sup>th</sup> amplification cycle. In these cases, the bacterial load was less than  $10^3$  GE/ml. Thus, the quantity of MOs needed to be at least  $10^3$  GE/ml for it to be considered above threshold, which meant that a positive signal was received in qPCR before the 35<sup>th</sup> cycle. The exceptions were *U. urealyticum*, *U. parvum*, *M. hominis* since there was no positive signal for these MOs in the negative control sample. If the signal was detected at any amplification cycle for these MO groups, qPCR result for them was regarded as positive. Yeast-like fungi of the *Candida spp.* were not included in this study.

### Statistical methods

The analysis of the structural characteristics of semen microbiota was carried out using the MSSC clustering model, which minimizes the sum over all clusters of intra-cluster sums of squared distances from cluster elements to their centroids [13]. The clustering problem was solved using the k-means++ algorithm [14], implemented in the scikit-learn machine learning library. The optimal clustering was selected on the basis of internal assessments of the clustering quality: the Silhouette index [15] and the Davies–Bouldin index (DBI) [16]. For optimal clustering, the stability of clusters to changes in the sample size was analyzed.

## RESULTS

### Detection rate for specific MO groups

TBL was detected in quantities exceeding the threshold value (at least  $10^3$  GE/ml) in 460 (72.5%) out of 634; samples the quantities of specific MO groups were below the threshold value in 110 (17.4%) of these 460 samples. Bacterial DNA was present in the quantities lower than  $10^3$  GE/ml in 174 (27.5%) samples.

From 1 to 14 MO groups were detected in quantities, exceeding the threshold value, simultaneously in 350 (55.2%) samples. Detection rate for specific MO groups is given in Table 1.

Different MO groups were found in a variety of associations with each other. Thus, we have decided to carry out cluster analysis in order to identify the microbial communities typical of semen microbiota.

### Cluster analysis of semen microbiota

For cluster analysis, 350 samples were selected in accordance with the following criteria: TBL in the quantity of at least  $10^3$  GE / ml, at least one group of MO in the quantity of at least  $10^3$  GE / ml.

To run the k-means++ clustering algorithm, each examined sample was represented as a vector  $(p, s) \in R^{50}$ , consisting of a vector of primary characteristics  $p \in R^{19}$ , (taken from the data on the semen microbiota analyses by means of qPCR) and secondary characteristics vector  $s \in R^{31}$ , calculated on the basis of primary characteristics.

The absolute values of the parameters determined by the Androflor kit (TBL and 18 MO groups) were regarded as primary characteristics.

The following secondary characteristics were calculated on the basis of the primary ones: corrected TBL (CTBL), equal to the total mass of the 18 MO groups detected by the kit; mass percentages of the MOs in relation to the CTBL; masses of the MO groups consolidated in accordance with the Androflor kit's configuration: lactobacilli, gram-positive facultative anaerobes (GPFA), obligate anaerobes (OA), gram-negative facultative anaerobes (GNFA), *Enterobacteriaceae / Enterococcus* (EE), and mycoplasmas; mass percentages of the consolidated MO groups in relation to the CTBL.

The optimal number of clusters in the examined dataset was determined on the basis of the values of the Silhouette and Davies-Bouldin indices (Table 2). The best clustering quality corresponds to the highest Silhouette Index and the lowest

Davies-Bouldin Index. In accordance with the obtained values of the indices, it was optimal to select 4 main clusters of the semen microbiota.

One consolidated MO group was predominant in each of the obtained clusters. The diagrams in Fig. 1 show the range of characteristics of the objects in their respective clusters.

Cluster 1 — the OA-dominated variant. The absolute quantity of all OA was comparable to the TBL and amounted to  $10^{4.3}$  GE / ml in the centroid (Fig. 1A). The proportion of OA in the centroid reached 82.8% in relation to the CTBL. This microbiota variant was identified in 172 (49.1%) out of 350 samples.

Cluster 2 — the lactobacilli-dominated variant. The absolute quantity of all lactobacilli was comparable to the CTBL and amounted to  $10^{4.0}$  GE / ml in the centroid (Fig. 1B). The proportion of lactobacilli in the centroid reached 80.9% in relation to the CTBL. This microbiota variant was identified in 78 (22.3%) out of 350 samples.

Cluster 3, characterized by the predominance of GPFA, was identified in 62 (17.7%) out of 350 samples. The absolute quantity of all GPFA was comparable to the CTBL and amounted to  $10^{3.6}$  GE / ml in the centroid (Fig. 1C). The proportion of GPFA in the centroid reached 89.4% in relation to the CTBL.

Cluster 4 — the EE-dominated variant. The absolute quantity of all EE was less than the CTBL and amounted to  $10^{3.5}$  GE / ml in the centroid (Fig. 1D). The proportion of EE in the centroid reached 64.5% in relation to the CTBL. This microbiota variant was identified in 38 (10.9%) out of 350 samples.

**Analysis of the microbial clusters' stability**

To analyze the stability of the identified clusters, subsamples of samples of  $f = 1,100$  volume of the original sample were generated (1000 random subsamples without return for each value of the volume). The generated subsamples were divided into 4 clusters. For each ( $m = 1,1000$ ) generated subsample of volume  $f$  samples (let us denote this subsample by  $X_m^f$ ) stability index of the  $k$  cluster was calculated using the following formula:

$$Stability\ Index(X_m^f, k) = \frac{\sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{true\}}([A(x_i), A(x_j) \in k] \wedge [\exists l A_m^l(x_i), A_m^l(x_j) \in l])}{\sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{true\}}([A(x_i), A(x_j) \in k]) \cdot \mathbf{1}_{\{true, false\}}([\exists l A_m^l(x_i), A_m^l(x_j) \in l])}, \quad (1)$$

where  $n$  is the number of samples in the subsample  $X_m^f$ ;  $\mathbf{1}_{\{true\}}: \{true, false\} \rightarrow \{0, 1\}$ ; is the indicator function of the logical argument;  $x \in X_m^f$  is the sample from the subsample  $X_m^f$ ,  $A(x)$ ;  $A^l(x)$  is the label of the cluster where the sample is contained as a result of clustering of the original sample set and the subsample  $X_m^f$ , respectively;  $k = \{1, 2, 3, 4\}$ ,  $l = \{1, 2, 3, 4\}$ , are the cluster labels.

**Table 2.** Clustering quality values

| Number of Clusters | Silhouette Index | Davies–Bouldin Index |
|--------------------|------------------|----------------------|
| 2                  | 0.21606          | 1.88844              |
| 3                  | 0.29118          | 1.55122              |
| 4                  | 0.32952          | 1.27021              |
| 5                  | 0.29094          | 1.642083             |
| 6                  | 0.30307          | 1.51617              |
| 7                  | 0.29392          | 1.58818              |
| 8                  | 0.29052          | 1.51856              |
| 9                  | 0.28759          | 1.43542              |
| 10                 | 0.29666          | 1.48578              |

In addition, the stability index of the  $k$  cluster, common for the subsamples of the  $f$  volume, was calculated. The calculations were carried out using the following formula:

$$Stability\ Index(k) = \frac{\sum_{m=1}^{1000} \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{true\}}([A(x_i), A(x_j) \in k] \wedge [\exists l A_m^l(x_i), A_m^l(x_j) \in l])}{\sum_{m=1}^{1000} \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}_{\{true\}}([A(x_i), A(x_j) \in k]) \cdot \mathbf{1}_{\{true, false\}}([\exists l A_m^l(x_i), A_m^l(x_j) \in l])}. \quad (2)$$

Fig. 2 shows the graphs depicting cluster stability indices calculated according to formulas (1) and (2). The obtained 4 clusters are stable: on sufficiently small volumes of subsamples, the probability of assigning two arbitrary observations to the same cluster with 4-clustering of the initial sample and an arbitrary subsample tends to 1. As follows from the graphs in Fig. 2, the most stable are the clusters with the predominance of lactobacilli (cluster 2, Fig. 2B), the predominance of GPFA (cluster 3, Fig. 2C) and with the predominance of EE (cluster 4, Fig. 2D). The least stable cluster is cluster 1 with the predominance of OA (Fig. 2A).

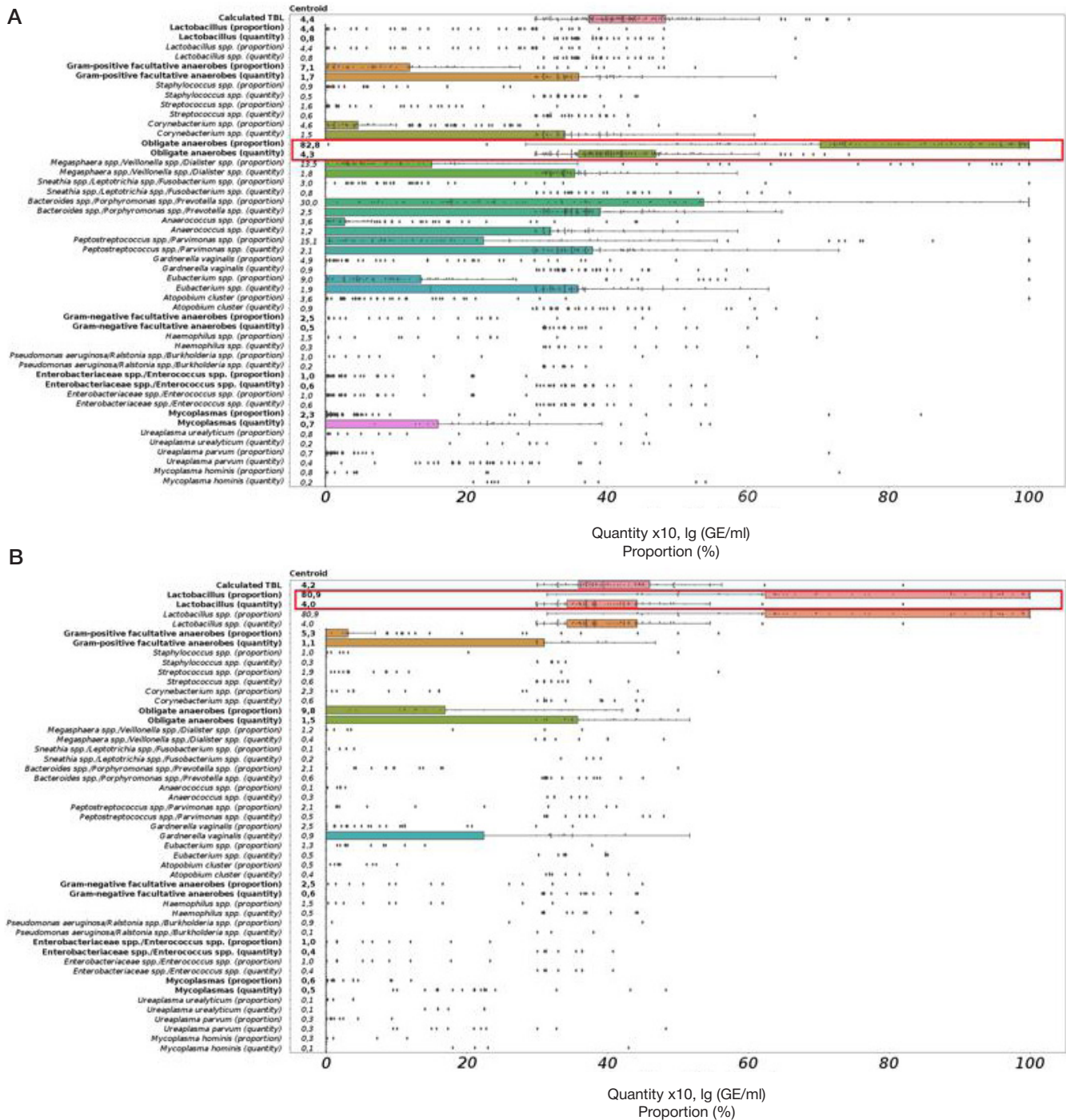
**DISCUSSION**

The presence of bacterial DNA both in the environment and in the reagents used for conducting the test (KITome) and high sensitivity of the PCR method limit our capability for interpreting results when analyzing the samples with low bacterial load [17]. Since positive signals were received for most MO groups after the 35<sup>th</sup> cycle in qPCR when analyzing negative control samples (which corresponded to the bacterial load of less than  $10^3$  GE/ml), the value of  $10^3$  GE/ml was regarded as the threshold value. All the other results were regarded as negative. The exceptions were *U. urealyticum*, *U. parvum*, *M. hominis* since there was no positive signal for these MOs in the negative control sample. If the signal was detected at any amplification cycle for these MO groups, qPCR result for them was regarded as positive.

The bacterial load in quantities exceeding the threshold value was identified only in 460 (72.5%) samples. The quantities of all the MO groups were below the threshold value in 110 of these 460 samples. Almost half of all the semen samples (44.8%) had bacterial DNA in the quantities below the threshold value (less than  $10^3$  GE/ml) which is regarded as a variant of the norm [12].

From 1 to 14 MO groups were detected simultaneously in quantities exceeding the threshold value in 350 (55.2%) samples, which corresponds with the results obtained by other researchers, who note that semen microbiota is heterogenous [1, 2, 5, 7]. The following MO groups were detected more often than others: *Corynebacterium spp.* (21,0%), *Bacteroides spp.* / *Porphyromonas spp.* / *Prevotella spp.* (20,7%), *Lactobacillus spp.* (19,7%), *Eubacterium spp.* (17,0%), *Peptostreptococcus spp.* /





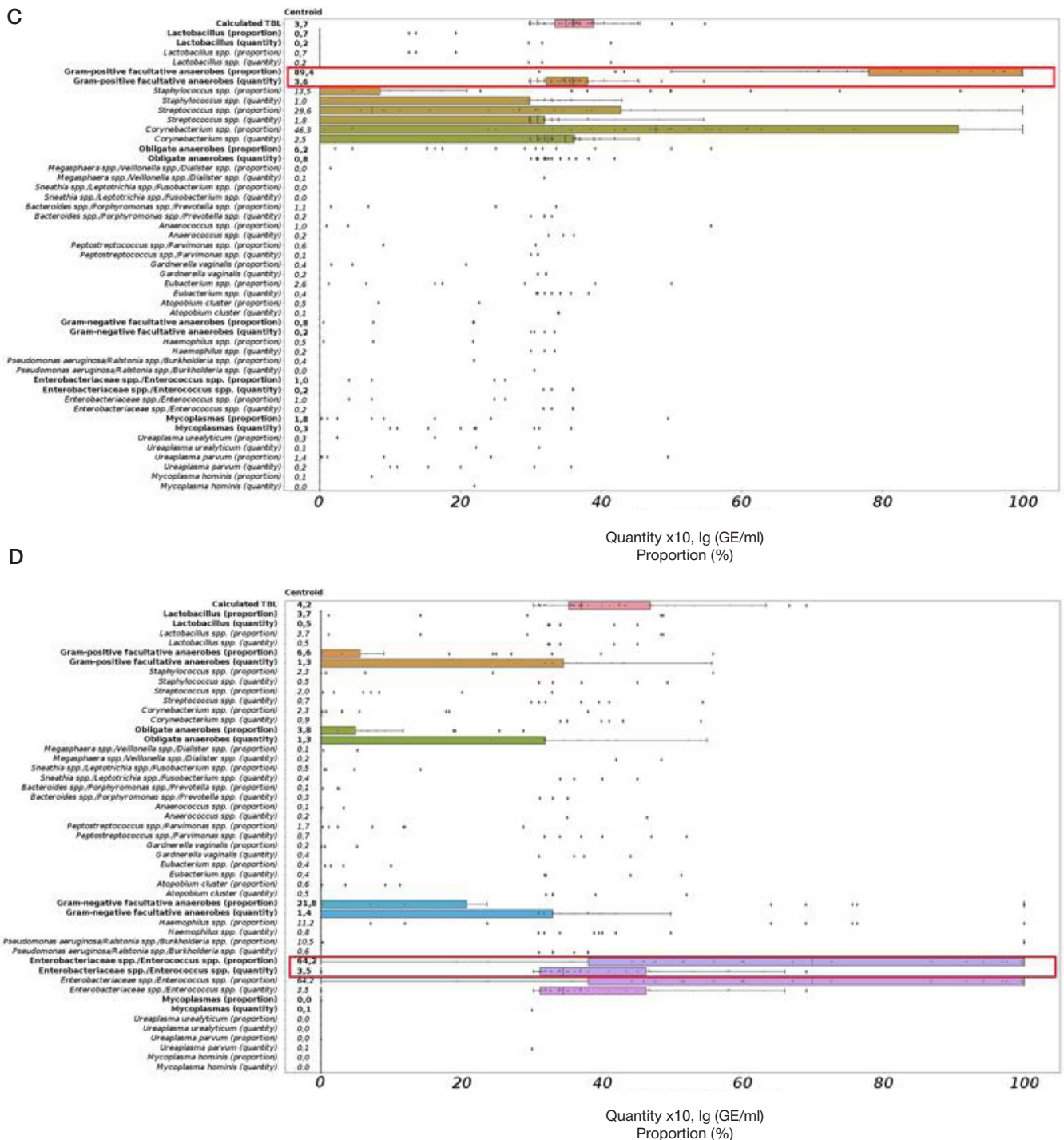
**Fig. 1.** Results of cluster analysis of semen microbiota analyzed by means of qPCR ( $n = 350$ ). The ordinate shows the values of the features in the centroid. Diagrams of the predominant groups of microorganisms are highlighted using red rectangles. Cluster 1 ( $n = 172$ ; Fig. **A**) is characterized by the predominance of obligate anaerobes, cluster 2 ( $n = 78$ ; Fig. **B**) is characterized by the predominance of lactobacilli, cluster 3 ( $n = 62$ ; Fig. **C**) is characterized by the predominance of gram-positive facultative anaerobes, cluster 4 ( $n = 38$ ; Fig. **D**) is characterized by the predominance of enterobacteria / enterococci

*Parvimonas* spp. (16,9%). Other MO groups were detected less often, with the rate of 3.3–14.8%. Previous studies have also shown that *Lactobacillus* spp. and obligate anaerobes along with facultative anaerobes and are often detected in the semen when using molecular-based methods [1, 2, 5, 7, 18].

Cluster analysis of semen microbiota in samples, containing TBL and at least one of the MO groups in quantities exceeding the threshold value, showed that division into 4 clusters was optimal. Each cluster was characterized by the predominance of one of the consolidated MO groups: cluster 1 — OA, cluster 2 — lactobacilli, cluster 3 — GPFA, cluster 4 — EE. Similar data were obtained in earlier studies using the NGS sequencing method

to evaluate the semen microbiota composition [1, 2]. Having studied the seminal fluid of healthy men and men with infertility, Hou D. et al. also identified several clusters of MOs, including those with predominance of GPFA, OA, and *Lactobacillus* spp. [2].

Clusters 2 (with the predominance of lactobacilli), 3 (with the predominance of GPFA), and 4 (with the predominance of EE) were characterized by high stability. Moreover, for clusters 2 and 3, the presence of other MO groups in the quantities comparable to those of the ones forming the cluster was atypical. At the same time, cluster 4 was characterized by the presence of other groups of bacteria, along with EE: GPFA, OA, and gram-negative facultative anaerobes.



**Fig. 1.** Results of cluster analysis of semen microbiota analyzed by means of qPCR ( $n = 350$ ). The ordinate shows the values of the features in the centroid. Diagrams of the predominant cluster of microorganisms are highlighted using red rectangles. Cluster 1 ( $n = 172$ ; Fig. A) is characterized by the predominance of obligate anaerobes, cluster 2 ( $n = 78$ ; Fig. B) is characterized by the predominance of lactobacilli, cluster 3 ( $n = 62$ ; Fig. C) is characterized by the predominance of gram-positive facultative anaerobes, cluster 4 ( $n = 38$ ; Fig. D) is characterized by the predominance of enterobacteria / enterococci

Cluster 1 (with the predominance of obligate anaerobes) was less stable. This may be due to the greater species diversity of the microbiota in these semen samples.

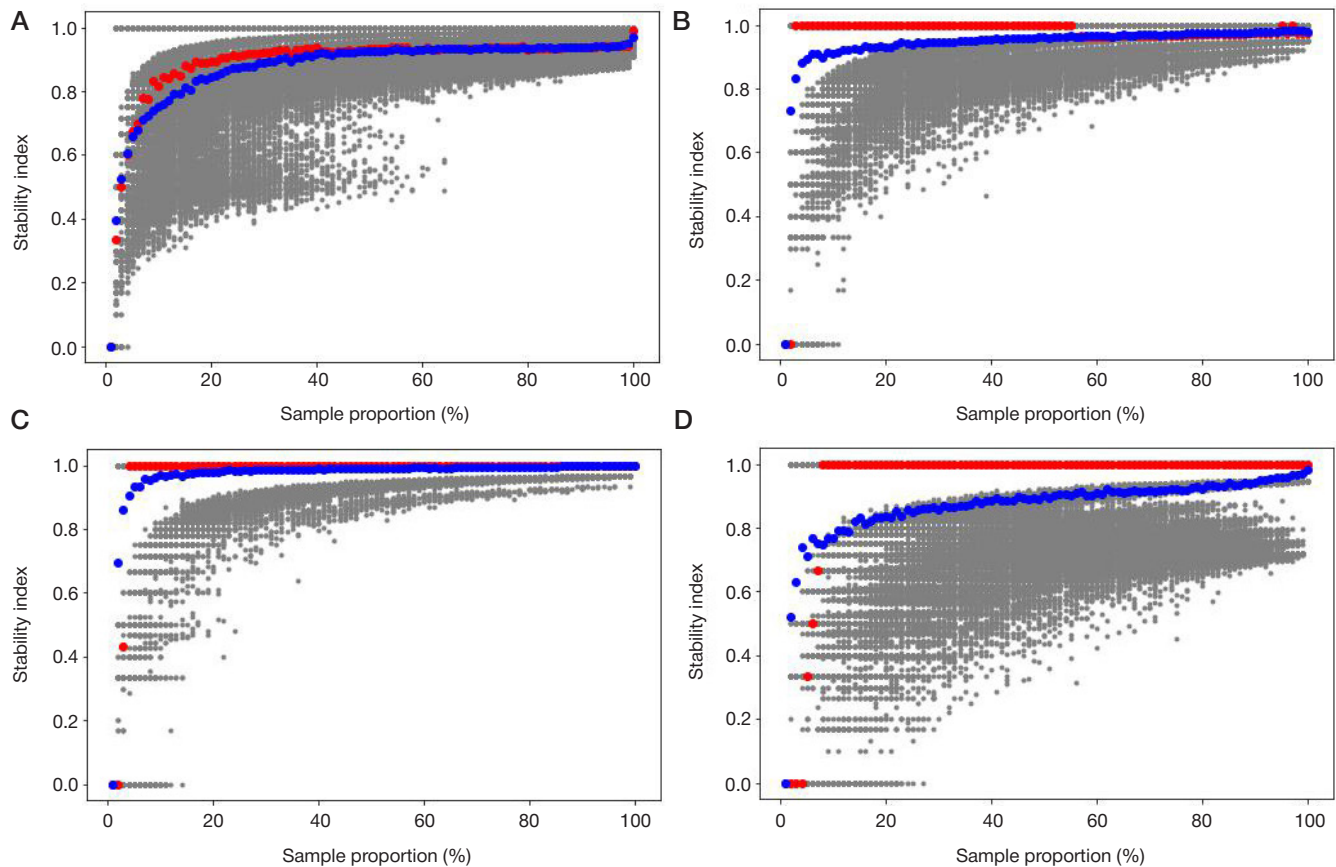
The results of this study confirm observations of other authors on the heterogeneous composition of the semen microbiota which can be grouped into a number of clusters. Our approach has confirmed the stability of the 4 clusters selected on randomly generated samples of different sizes.

Further research is necessary to determine the detection rate of the described bacterial clusters in semen with normospermia and various types of pathospermia. We need

to establish the relationship between the characteristics of the semen microbiota and infertility in men. This will allow the development of new algorithms for treating patients with reproductive disorders, depending on the composition of the semen microbiota.

**Conclusion**

1. Bacterial DNA was detected in the quantity of at least  $10^3$  GE/ml in 72.5% of the semen samples by means of qPCR; in 55.2% semen samples, from 1 to 14 MO groups were detected in



**Fig. 2.** Results of the cluster stability analysis 1 (A), 2 (B), 3 (C), 4 (D). The grey marker on the graphs shows cluster stability index on a random subsample, the red marker shows the median of the stability indices, calculated for 1000 random subsamples of the  $f$  volume, the blue marker shows the cluster stability index on a set of the  $f$  volume

quantities exceeding the threshold value. 2. We identified 4 stable clusters of semen microbiota. A certain consolidated MO group was predominant in each of these clusters: obligate anaerobes, lactobacilli, gram-positive facultative anaerobes,

*Enterobacteriaceae* / *Enterococcus* group. 3. In half of the samples microbiota was represented by cluster 1 (with obligate anaerobes being the predominant group), which was the least stable one and was characterized by the greatest species diversity.

## References

- Baud D, Pattaroni C, Vulliamoz N, Castella V, Marsland BJ, Stojanov M. Sperm Microbiota and Its Impact on Semen Parameters. *Front Microbiol.* 2019; 10: 234. DOI: 10.3389/fmicb.2019.00234.
- Hou D, 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): 1261–9. DOI: 10.1016/j.fertnstert.2013.07.1991.
- Liu CM, Osborne BJ, Hungate BA, Liu CM, Osborne BJ, Hungate BA, et al. The semen microbiome and its relationship with local immunology and viral load in HIV infection. *PLoS Pathog.* 2014; 10 (7): e1004262. DOI: 10.1371/journal.ppat.1004262.
- Mändar R, Punab M, Korrovits P, Türk S, Ausmees K, Lapp E, et al. Seminal microbiome in men with and without prostatitis. *Int J Urol.* 2017; 24 (3): 211–216. DOI: 10.1111/iju.13286.
- 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): e110152. DOI: 10.1371/journal.pone.0110152.
- Zozaya M, Ferris MJ, Siren JD, Lillis R, Myers L, Nsuami MJ, et al. Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. *Microbiome.* 2016; 4: 16. DOI: 10.1186/s40168-016-0161-6.
- Štšepetova J, Baranova J, Simm J, Parm Ü, Rööp T, Sokmann S, et al. The complex microbiome from native semen to embryo culture environment in human in vitro fertilization procedure. *Reprod Biol Endocrinol.* 2020; 18 (1): 3. DOI: 10.1186/s12958-019-0562-z.
- Kiessling AA, Desmarais BM, Yin HZ, Loverde J, Eyre RC. Detection and identification of bacterial DNA in semen. *Fertil Steril.* 2008; 90 (5): 1744–56. DOI: 10.1016/j.fertnstert.2007.08.083.
- Pochernikov DG, Galkina IS, Postovoytenko NT, Gerasimov AM. A comparative analysis of seminal and vaginal microbiota of married couples by real-time PCR with Androflor and Femoflor reagent kits. *Bulletin of Russian State Medical University.* 2017; 2: 34–39.
- Voroshilina ES, Zornikov DL, Panacheva EA. Evaluation of the ejaculate microbiota by real-time PCR and culture-based technique. *Bulletin of Russian State Medical University.* 2019; 1: 41–5.
- Pochernikov DG, Vitvickaja JuG, Boldyreva MN, Galkina IS. Informativnost' biomateriala dlja issledovanija mikrobioty urogenital'nogo trakta muzhchin metodom PCR-RV (pilotnoe issledovanie). *Jeksperimental'naja i klinicheskaja urologija.* 2019; 2: 128–33. Russian.
- Schuppe HC, Pilatz A, Hossain H, Diemer T, Wagenlehner F, Weidner W. Urogenital Infection as a Risk Factor for Male Infertility. *Dtsch Arztebl Int.* 2017; 114 (19): 339–46.
- Jain AK, Murty MN, Flynn PJ. Data clustering: a review. *ACM computing surveys (CSUR).* 1999; 31 (3): 264–323.
- Arthur D, Vassilvitskii S. K-means++: the advantages of careful seeding. In: Hal Gabow, editor. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*



- (Proceedings in Applied Mathematics) 18th ed. Society for Industrial and Applied Mathematics (3 January 2007). P. 1027–1035.
15. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987; 20: 53–65.
  16. Davies DL, Bouldin DW. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* PAMI-1. 1979; (2): 224–27.
  17. de Goffau MC, Lager S, Salter SJ, Wagner J, Kronbichler A, Charnock-Jones DS, et al. Recognizing the reagent microbiome. *Nat Microbiol.* 2018 Aug; 3 (8): 851–3.
  18. Yang H, Zhang J, Xue Z, Zhao C, Lei L, Wen Y, et al. Potential Pathogenic Bacteria in Seminal Microbiota of Patients with Different Types of Dysspermatism. *Sci Rep.* 2020; 10: 6876.

## Литература

1. Baud D, Pattaroni C, Vulliamoz N, Castella V, Marsland BJ, Stojanov M. Sperm Microbiota and Its Impact on Semen Parameters. *Front Microbiol.* 2019; 10: 234. DOI: 10.3389/fmicb.2019.00234.
2. Hou D, 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): 1261–9. DOI: 10.1016/j.fertnstert.2013.07.1991.
3. Liu CM, Osborne BJ, Hungate BA, Liu CM, Osborne BJ, Hungate BA, et al. The semen microbiome and its relationship with local immunology and viral load in HIV infection. *PLoS Pathog.* 2014; 10 (7): e1004262. DOI: 10.1371/journal.ppat.1004262.
4. Mändar R, Punab M, Korrovits P, Türk S, Ausmees K, Lapp E, et al. Seminal microbiome in men with and without prostatitis. *Int J Urol.* 2017; 24 (3): 211–216. DOI: 10.1111/iju.13286.
5. 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): e110152. DOI: 10.1371/journal.pone.0110152.
6. Zozaya M, Ferris MJ, Siren JD, Lillis R, Myers L, Nsuami MJ, et al. Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. *Microbiome.* 2016; 4: 16. DOI: 10.1186/s40168-016-0161-6.
7. Štšepetova J, Baranova J, Simm J, Parm Ü, Rööp T, Sokmann S, et al. The complex microbiome from native semen to embryo culture environment in human in vitro fertilization procedure. *Reprod Biol Endocrinol.* 2020; 18 (1): 3. DOI: 10.1186/s12958-019-0562-z.
8. Kiessling AA, Desmarais BM, Yin HZ, Loverde J, Eyre RC. Detection and identification of bacterial DNA in semen. *Fertil Steril.* 2008; 90 (5): 1744–56. DOI: 10.1016/j.fertnstert.2007.08.083.
9. Почерников Д. Г., Галкина И. С., Постовойтенко Н. Т., Герасимов А. М. Сравнительный анализ биотопа эякулята и цервикального канала методом ПЦР-РВ с тестами «Андрофлор» и «Фемофлор» в супружеских парах. *Вестник Российского государственного медицинского университета.* 2017; 2: 37–41.
10. Ворошилина Е. С., Зорников Д. Л., Паначева Е. А. Сравнительное исследование микробиоты эякулята методом количественной ПЦР и культуральным методом. *Вестник Российского государственного медицинского университета.* 2019; 1: 44–9.
11. Почерников Д. Г., Витвицкая Ю. Г., Болдырева М. Н., Галкина И. С. Информативность биоматериала для исследования микробиоты урогенитального тракта мужчин методом ПЦР-РВ (пилотное исследование). *Экспериментальная и клиническая урология.* 2019; 2: 128–33.
12. Schuppe HC, Pilatz A, Hossain H, Diemer T, Wagenlehner F, Weidner W. Urogenital Infection as a Risk Factor for Male Infertility. *Dtsch Arztebl Int.* 2017; 114 (19): 339–46.
13. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. *ACM computing surveys (CSUR).* 1999; 31 (3): 264–323.
14. Arthur D, Vassilvitskii S. K-means++: the advantages of careful seeding. In: Hal Gabow, editor. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Proceedings in Applied Mathematics) 18th ed.* Society for Industrial and Applied Mathematics (3 January 2007). P. 1027–1035.
15. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987; 20: 53–65.
16. Davies DL, Bouldin DW. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* PAMI-1. 1979; (2): 224–27.
17. de Goffau MC, Lager S, Salter SJ, Wagner J, Kronbichler A, Charnock-Jones DS, et al. Recognizing the reagent microbiome. *Nat Microbiol.* 2018 Aug; 3 (8): 851–3.
18. Yang H, Zhang J, Xue Z, Zhao C, Lei L, Wen Y, et al. Potential Pathogenic Bacteria in Seminal Microbiota of Patients with Different Types of Dysspermatism. *Sci Rep.* 2020; 10: 6876.