Metagenome Profiling of Anterior Uterine Samples to Improve in-vitro Fertilization Success Ratio

Gulab Khedkar (gdkhedkar@gmail.com)
Dr Babasaheb Ambedkar Marathwada University https://orcid.org/0000-0002-2060-785X
Manju Jilla
Dr Babasaheb Ambedkar Marathwada University
Vidya Niranjan
R V College of Engineering
CD Khedkar
College of Dairy Sciences
Amol Kalyankar
Dr Babasaheb Ambedkar Marathwada University
Rajshree Deolalikar
Dr Babasaheb Ambedkar Marathwada University
Chaitali Khedkar
SBMT Medical college
Bharathi Prakash
university college, Mangalore
P Josthna
Sri Padmavathi mahila university
Pramod Yeole
Dr Babasaheb Ambedkar Marathwada University
K Piyusha
RV College of Engineering
C Lavanya
RV College of Engineering

Research

Keywords: In-Vitro Fertilization, metagenome, cervix, vagina, infertility, assisted reproductive technology

Posted Date: November 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1020192/v1

Abstract

Background: The radical increase in the number of couples experiencing infertility, which is the inability to conceive in a natural manner, has increased the interest among the researchers worldwide to identify the root cause for this problem. Relevant studies are needed to ascertain the cause behind unexplained infertility, and certain interventions are needed in the process of assisted reproductive technology to improve the implantation success. The presence of certain pathogenic flora in the cervix and vaginal region are often attributed to the failure of In-vitro fertilization (IVF), which is a method of assisted reproduction. This study explored the cervix and vaginal microbiota of 52 such infertile females who underwent assisted reproduction (IVF), to analyse the microbiota, identify pathogenic flora and evaluate the diversity of the flora.

Methods: 52 microbial swab samples of infertile females were sequenced using 16s rRNA approach. 16s rRNA data analysis was performed to get the taxonomic abundance and pathogenicity of the resultant microbiota. The microbial communities were statistically assessed for probable similarities and differences between positive, negative and not applicable outcomes of IVF.

Results: A total of 31 genera were identified across the study samples, and a thorough investigation of associated pathogenicity of these taxa was performed. Certain pathogenic genera specific to positive as well as negative implantation were observed. Additionally, the alpha diversity was observed to be higher in the cervical region compared to the vaginal region. Rarefaction analysis was performed. The analysis of beta diversity was performed and the phylogenetic distance was visualized through distance boxplots.

Conclusion: The presence of pathogenic bacteria, especially those associated with conditions like Bacterial Vaginosis and Chronic Endometritis, are often linked to infertility conditions and were predominant in the cervical and vaginal samples. Dysbiosis in the cervical and vaginal region is often attributed to decrease in *Lactobacillus spp.* and increase in other species, leading to difference in functionalities and growth of pathogens. In this study, the presence of pathogenic organisms in both regions was remarkable. Additionally, large bacterial diversity in cervical regions was observed, which might be a possible reason for implantation failure. Further studies on the mechanism of action of these bacteria is needed; the possible sources of contamination, such as the tip of the catheter during embryo transfer, the DNA extraction kits prior to sequencing, etc. need to be identified as this might introduce pathogens during the procedure. By using more sterility in the procedure as well as profiling the metagenome of the infertile female, novel insights can be gained as well as a more directed procedure can be performed to increase the chances of success in implantation.

Background

Infertility is a condition, also sometimes referred to as a disease, generally characterized by the failure to achieve clinical pregnancy even after unprotected sexual intercourse. A variety of factors are generally

considered to be associated with infertility. Some of the common factors that are generally considered are the age of the female, the medical history of the male and female, the reproductive health, etc [1].

Infertility affects up to 15% of couples in their reproductive years around the world. The overall occurrence of primary infertility in India, according to the World Health Organization, ranges from 3.9 to 16.8% [2]. In some cases, couples experience infertility without any explainable clinical background/history. When such patients resort to assisted reproductive techniques like *In-vitro* fertilization (IVF), a large proportion of couple's experience implantation failure. Recurrent Implantation failure is usually observed in women aged less than 40 years, where even after four embryos (high quality) transfers in a minimum of three fresh or frozen cycles, they do not achieve pregnancy.

It qualifies as multiple unsuccessful implantation/ IVF cycles. One failed IVF attempt has multiple disadvantages for a couple which include health of the female, psychological pressure and also the financial compromises. The major factors attributed to the condition of infertility common to both the genders are hormonal disorders, sexually transmitted disorders, lifestyle disorders such as smoking, alcohol consumption, obesity, etc. [3].

The female reproductive tract consists of the uterus, cervix and the vaginal region and is home to a diverse community of bacteria, though low in abundance [4]. Several studies have pointed to the finding that the female reproductive tract has an active and healthy state of microbiota when it is inhabited by bacteria of the *Lactobacillus* genus. A variety of factors influence the microbiota of the cervicovaginal area, namely menstrual cycle, sexual behaviour, state of infection etc. [5].

In the current study, an attempt was made to discover the microbiota of infertile females by conducting a 16s microbial assay on 52 infertile females. A variety of parameters such as the age, number of infertility years, result of IVF treatment etc., were noted. The comparison of metagenome of the cervix and the vagina of infertile females had been made to assess the changes occurring due to certain factors. An effective approach of assessing the problem was to profile the metagenome of infertile women, of both the cervical and vaginal region, to observe the differences in microbiota. The identification of pathogenic flora in these regions can be assessed and the probable modes, internal or external, in which these microbes inhabit the anterior uterine region, can be identified.

This study addresses the problem of infertility and tries to identify the possible relation between the microbiota and infertility condition. It is also aimed at identifying the functional characteristics of the microbial flora and pathogenicity, along with the possible sources of the same. The study also analyses the nature of microbial habitat in cervix/vaginal region and its possible effect on the outcome of assisted reproductive techniques like IVF.

The objectives of this study were mainly to qualitatively and quantitatively analyse the microbial community associated with cervix and vaginal samples, identify the most prevalent bacteria in cervix and vaginal samples along with their associated pathogenicity analysis of the microbiota.

Methodology

SAMPLING, COLLECTION AND EXPERIMENTATION

Initially, a written informed consent was acquired from infertile patients having a history of Recurrent Implantation Failure (RIF), all in accordance with ICMR declaration. Those patients undergoing the primary IVF attempt in high-volume centres with greater than 2000 oocyte pickup cycles per year, were also enrolled into the study. A total number of 52 patients were recruited for the study. The participants were provided both pictorial and verbal instructions that directed them to provide swabs in a private examination room. Swabs for microbial sequencing were carefully collected and sampled in sterile Catch-A11^{TM} sample collection swabs (Epicentre Biotechnologies, Madison, WI). These were instantly plunged into MoBio bead tubes (MoBio Laboratories, Inc., Carlsbad, CA) and frozen on dry ice in an upright position until they were moved to the lab and stored at -80^{o}C for further processing.

Infertility inspections such as hysteroscopy, hysteronsal pingogram, thyroid functions and immunological and thrombophilic factor examinations were thoroughly conducted. Morphological classification was done according to Veek's classification [6]. Levels of serum human chorionic gonadotropin (hCG, Tosoh Co., Shunan, Japan) was measured on the 11th day post transfer of day 3 embryos or on the 9th day, after day 5 blastocysts transfer. As per the instructions provided by the manufacturer, a value lesser than 0.5 IU/L is regarded as negative for a pregnancy test. The RIF values was thus defined as serial negative pregnancy tests that follow that transfer of 5 or more cleavage stage embryos and/or blastocysts that were found to be morphologically good (Supplementary Table S1).

Microbial DNA isolation

Microbial DNA for the samples were isolated using the QIAamp DNA microbiome kit (Qiagen, Hilden, Germany). Adequate amplification of the 16s rRNA V3-V4 region was recovered after the initial PCR run of the full-length 16S rRNA genes. The primer pairs used were 5'- AGAGTTTGATCMTGGCTCAG–3' for forward and 5'- TACGGYTACCTTGTTACGACTT-3' for reverse. The PCR was run for 25 cycles. The amplicon obtained as a result of PCR amplification was then utilised as the template for the amplification of the V4 region using the primers 5'- GTGCCAGCMGCCGCGGTAA -3' and 5'- GGACTACHVGGGTWTCTAAT -3') for forward and reverse respectively for 30 cycles in triplicates. The PCR reaction of 25 µL comprises 2x KAPA HiFi HotStart ReadyMix (12.5 µL), PCR primers each (1µM), microbial genomic DNA (5 ng/µL).

Quality Check

The read-wise assessment of 16s rRNA sequencing quality in both cervical and vaginal samples was performed using FASTQC [7].

Taxonomy Assignment

Taxonomical assignment was done using two different databases mentioned below for three experimental conditions namely Positive implantation, Negative implantation, Not Applicable (NA) implantation. Positive implantation refers to the successful pregnancy after 15 days of embryo transfer and negative implantation indicates the inability to get conceived after 15 days of embryo transfer. Not applicable (null) implantation refers to inability to transfer embryos due to genetic abnormalities.

RDP DATABASE

The Ribosomal Database Project (RDP) is an annotated and aligned collection of ribosomal RNA sequence data. It also hosts several tools, enabling scientists worldwide for analysis of their own rRNA data within the framework of RDP. It contains more than 2800000 annotated and aligned archaeal and bacterial rRNA sequences and greater than 62000 fungal rRNA gene sequences. The RDP Pipeline tools have capacity to handle high volumes of amplicon data [8].

It contains sequences from International Sequence Database Collaboration (INSDC). The taxonomic classification stands on the basis of Bergey's Trust and LPSN [9].

The pre-processing and analysis (merging of reads, filtering, barcoding, dereplication, abundance sorting, singleton discarding, OTU Clustering, OTU table creation and calculation of alpha diversity) of the reads were carried out using USEARCH [10]. Analysis of microbial abundance and visualization of reads was performed using the Phyloseq package in R programming language [11]. The plots were generated for three different groups of experiment namely, Positive, Negative and not applicable implantation.

GREENGENES DATABASE

The Green genes database is also designed for archaea and bacteria, and performs *de-novo* building of trees and mapping of rank from other sources of taxonomy. The 16s rRNA sequences are fetched from public databases and processed subsequently, construction of phylogenetic tree is carried out. Alignment is based on sequence and structure and Fast Tree performs tree construction. As an inbuilt database the QIIME tool uses green genes latest updated in 2013 [9].

Green genes offer four specific features namely, ARB compatibility, a set of descriptive fields, assignment of taxonomy and screening of chimeras. Chimera is sequences which are formed due to a combination of two or more biological sequences. Additionally, 3 types of non-multiple sequence alignment matches are available, namely, extension of SEED by BLAST, shared 7-mer similarity (Simrank) and degenerative pattern match for probe/ primer valuation [12].

The cervix and vaginal reads were analysed using QIIME1.9.1 [13] (merging of reads was performed using FLASH Software [14], OTU picking, alpha diversity calculation, Rarefaction curve, Beta diversity calculation, Principal coordinate analysis was performed in QIIME1.9.1)

Results And Discussion

Sample collection and processing

As per Veeck's classification [6], cleavage-stage embryos that were morphologically good were defined as day 3, grade 1 or 1, seven-to-nine cell embryos. Further, according to Gardner's score [15], morphologically good blastocysts were found to have a score of 3BB and above, and were defined as day 5 blastocysts.

The products obtained as a consequence of PCR were quantified using a Qubit fluorimeter (Invitrogen, Carlsbad, CA) and run on a bioanalyser to confirm the appearance of a single band at 350bp. Furthermore, no band was to be observed for the "kit-ome" control.

QUALITY CHECK

The supplementary figure S1 depicts the consolidated quality check result of all reads contained in the 52 samples. The different FASTQC check parameters are basic statistics, per base sequence quality, per tile sequence quality, per sequence quality score, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication level, overrepresented sequences, adapter content, k-mer content.

TAXONOMY ASSIGNMENT

RDP Database

Cervical Samples

Due to a large number of species recognized, subsequent filtering was done based on abundance of the microbes within the sample. The taxa were filtered in such a way that only those OTUs were considered which represented at least 20 percent of reads in at least one sample. This filtering was done and the graph was generated at the genus level taxonomy.

Among the taxa obtained at genus level for the cervix samples, the maximum abundance is seen for *Lactobacillus* genus. The top genera observed were *Alcanivorax, Bifidobacterium, Corynebacterium, Defluviitoga, Enterococcus, Escherichia/Shigella, Gardnerella, Geodermatophilus, Lactobacillus, Marininema, Neorhizobium, Owenweeksia, Parcubacteria_genera_incertae_sedis, Prevotella, Proteus, Staphylococcus, Streptococcus, Streptophyta.* The barplot for the abundance values in cervical samples is shown in figure 1.

Vaginal samples

As described elsewhere, the filtering of taxa was done to obtain the genera that represent at least 20 per cent of the reads in every sample.

Here, the taxa from Negative controls were filtered off (figure 1A) and the resulting dominant taxa in vaginal samples according to RDP database are: *Alcanivorax, Bifidobacterium, Elizabethkingia, Enterococcus, Escherichia/Shigella, Gardnerella, Lactobacillus, Neorhizobium, Pantoea, Staphylococcus,*

Streptococcus. The barplot for the abundance values in cervical samples is shown in figure 1B. The species level analysis from the RDP database for both cervix and vaginal samples is given as **supplementary figure S2 and S3**.

GREENGENES DATABASE

Another prominent database commonly employed in taxonomic annotation to read is the Greengenes database. It, however, reports up to the level of genus for a given sample. The graphs were generated separately for each implantation type.

The above figure 2 depicts all the possible genera obtained in each sample in comparison against Green genes database. Since a large number of organisms, filtering based on abundance was carried out and only the taxa with abundance greater than 10 per cent in a sample have been reported.

Cervix - positive implantation results

Overall, the abundant genera in cervix representing positive implantation are: *Lactobacillus, Bifidobacterium, Alcanivorax, Acinetobacter, Enterobacteriaceae, Rhizobiaceae, Enterococcus, Haererehalobacter, Pelagibacteraceae, Gardnerella, Streptococcus, Ralstonia, Ureaplasma, Rhodobacteraceae*. The barplot indicating the microbes in positive implantation and the abundance is shown in figure 2A.

Cervix - negative implantation results

Filtering of the genera based on its abundance in the sample (> 10%) was performed to yield the following abundant genera: *Haererehalobacter, Alcanivorax, Lactobacillus, Enterococcus, Enterobacteriaceae, Staphylococcus, Proteus, Bifidobacterium, Lactococcus, Streptococcus, Haemophilus.* The barplot indicating the microbes in negative implantation and the abundance is shown in figure 2B.

Vagina - Positive implantation results

The genera obtained in the vaginal samples of the infertile females are shown above. According to the filtered results, the abundant genera are: *Lactobacillus, Bifidobacterium, Enterobacteriaceae, Enterococcus, Elizabethkingia, Ralstonia, Alcanivorax, Ureaplasma, Gardnerella, Streptococcus, Acinetobacter* (Figure 3A).

Vagina - Negative implantation results

The result obtained for negative or unsuccessful implantation is shown above. The filtered abundant genera are: *Lactobacillus, Bifidobacterium, Enterobacteriaceae, Enterococcus, Prevotella, Staphylococcus, Planococcaceae, Proteus, Alcanivorax* (Figure 3B).

Microbial communities obtained having null effect on implantation success

Cervix – NA implantation results

The most abundant genera obtained in the NULL category are: *Lactobacillus, Gardnerella, Ureaplasma, Haererehalobacter, Alcanivorax, Enterococcus.* The barplot indicating the microbes in NA implantation and the abundance is shown in figure 4A.

Vagina – NA implantation results

The result obtained for null implantation is shown above. The filtered abundant genera are: *Lactobacillus*, *Gardnerella, Ureaplasma, Alcanivorax, Enterococcus* (Figure 4B).

Pathogenicity report of the present microbial community

The table 1 shows the list of all the genera present in all three experimental conditions (Positive, negative, not applicable) highlighting the unique and the specific genera with respect to negative outcome. The details about the pathogenicity of each of genera is given in Supplementary table S2.

Statistical Analysis

The statistical analysis of 16s rRNA data generally involves computation of alpha diversity (measure of diversity contained within a sample) and beta diversity (diversity between the samples).

Alpha Diversity

The Green genes database reports diversity till genus level of taxonomy and RDP database reports till species level of taxonomy.

Cervical Samples

The intra-sample diversity measure in the form of a violin plot is depicted in figure 5. A violin plot is a combination of a boxplot and a histogram. Here, the chosen index for representation is the Shannon's Entropy index. Shannon Diversity index is a mathematical measure of species diversity in a community. The Shannon Diversity index also takes into account both the number of species present in a community and their relative abundance [15]. The measurement of diversity consists of two components: 1) Species Richness: these depict the number of individuals per sample: 2) Equitability Index: the different species in a community are compared in terms of evenness of their distribution; the relative abundance is measured.

A community with a high value of evenness index implies that it has the same number of individuals in different species. The number of species present in the community is said to be species richness index. For a community consisting of a single species, the least value of Shannon Index (H') is zero. There is maximum content of information in the sample if there is high species evenness (equal number in all species) [16]. This index increases as the species richness and evenness both increases. The Shannon index also assumes that all the species are represented in the sample. Typical values of Shannon Index

lie between 1.5 and 3.5 which is in confirmation to our observations of this study as shown in figure 5(Aa and Ab).

Observing the violin plots for both cervical and vaginal samples (from Green genes database), we can see that the largest value of Shannon Index is quite high compared to the limits (i.e., nearly 8 in cervix and up to 6 in vagina). In the cervix samples, it can be seen that the highest Shannon entropy is observed in the set of samples whose implantation result is positive (Figure 5Aa). The next highest Shannon Entropy is that of Negative implantation result (Figure 5Aa). The Shannon index values of vaginal samples are very low compared to the cervical samples (Figure 5Ba and Bb). The values of the index can be deduced from the inner box plot situated within the symmetrical histogram. The higher range of values in cervical samples implies higher species richness and evenness. The vaginal samples have lower species richness. But, on the other hand, if we compare the density plots surrounding the boxplot in both cervix and vagina, we can see that a majority of the values are clustered in a single place for the vaginal samples. For positive samples in the vagina, the curve almost resembles a normal distribution (Figure 5Ab and Bb). In the cervical samples however, the frequency of index values is not very high, but the range of values is quite wide (Figure 5Aa and Ba).

The plots indicate alpha diversity from the RDP database report at the species level of classification. The range of diversity depicted by the RDP database seems to be lower compared to that of the Green genes database. Additionally, the species level alpha diversity also reports overall higher diversity in the cervix samples compared to vaginal samples (Figure 5). A large proportion of the Shannon Index values are clustered within 0-1 range for the RDP database plots.

To summarize, at the genus level, Green genes database reports a high Index value depicting large species richness and evenness (with larger values in cervix compared to vagina). From the RDP database plots (at species level), it can be deduced that the diversity till species level is less. The range of alpha values are lesser compared to Green genes. This indicates that there is a wide variety of genera but a comparatively lower diversity of species among the samples.

Rarefaction Analysis

Generally, rarefaction analysis is performed to identify the species richness by the way of sampling. This is identified using the concept of rarefaction curve, which is a curve of the number of species against the number of samples. As the initial sampling is being done at the commencement of the analysis, the curve steadily rises as the most common species are rapidly being identified. However, once this is done, the curve reaches a plateau as only the rarest species are yet to be identified. The downfall of this method is that if the size of subjects being sampled keeps increasing, the number of species being identified also keeps increasing. Thus, the method of repeatedly re-sampling a given number of samples is done and the number (average) of species identified in each sample is plotted.

Three estimators are used for alpha rarefaction: chao1, observed OTUs and PD (Phylogenetic diversity) whole tree. Rarefaction is a method for creating an even number of reads for a given sample. This is

crucial because during comparison of diversity, higher numbers of samples reveal rare species. Additionally, when similarities of samples are considered, there might be findings of rare species in a sample because of higher sampling depth but not in another sample. Thus, rarefaction is done to account for differences in sampling depth while comparison for similarities or differences between samples.

Rarefaction curves, however, are different from rarefaction. Rarefaction accounts for uneven/ unequal sampling size whereas rarefaction curves identify if all the diversity present in a given community is captured. The curve between the number of reads and the number of OTUs is the rarefaction curve. The flattening of the curve depicts that the sample had enough reads to identify most of the diversity contained [17].

Chao1 is a method of estimation of species richness (number of different species present in an environment) based on abundance. It is a non-parametric technique, and gives more relevance to species which have less abundance (only doubletons and singletons give the number of missing species). It assumes that rare species give most information about the number of missing species [18]. The PD whole tree is a measure of diversity which identifies, among a set of species, the degree of relatedness (in the form of a tree). It works on the basis that the diversity of a set of species is less when they are closely related and more when they are distantly related [19]. The main viewpoint of analysis of the rarefaction curve is that if the curve for any given category do not extends all the way to the right end of the X-axis, then it means that at minimum one of the samples in that category does not have enough sequences. The rarefaction analysis was conducted for both the cervical and vaginal samples (Figure 6). Since all the plots could not be displayed, alpha rarefaction curves for some of the important categories are shown.

Cervical Samples

Here, three curves have been plotted for positive result of implantation: the metadata variables considered are age (Figure 6A), assisted hatching (Figure 6B) and number of embryos transferred (Figure 6C). It can be seen that all the three curves have nearly flattened curves indicating that enough sequences per sample were present to capture the diversity. For the number of embryos transferred (Figure 6B), one curve hasn‘t completely reached the plateau.

Vaginal Samples

Amongst the vaginal samples, the alpha rarefaction curve for the following metadata categories has been included for positive result of implantation (Figure 7A), procedure opted for the assisted reproductive treatment (figure 7B) and cause of infertility and type of infertility (primary or secondary) (Figure 7C). The curve for type of procedure opted and Chao metric has not reached plateau indicating lack of sequencing depth in identification of species richness.

Beta Diversity and principal coordinate analysis

The beta diversity is calculated to identify the diversity between samples. Here, the Weighted and Unweighted Unifrac distance measure was chosen to find if there is any significant clustering between the samples of three outcome categories, namely Positive, Negative and Null. Since the beta diversity measures are dissimilarity measures, higher numbers indicate lesser similarity and more differences between the samples/ communities under consideration.

Since the Unifrac metric is being used here as the distance measure to calculate the pairwise distances from each sample to all other samples, it results in a large distance matrix. This distance matrix is then made to undergo the ordination process, where the matrix is converted into all samples and the respective principal axes of variation in a process called ordination. Generally, three principal components are measured for easy visualization. The principal coordinate analysis (PCoA) is used to represent the ordinated matrix in the form of the first three principal axes of variation to observe any clustering based on the required outcome or metadata variable [20]. Unifrac takes into consideration the phylogenetic information for comparison of samples. The main contributing factors responsible for the variation between the microbial communities can be identified through PCoA [21]. There are two kinds of Unifrac metrics, namely weighted, also known as quantitative, which takes abundance of organisms into consideration and unweighted, which takes only its presence or absence and hence is called qualitative in nature.

Cervical Samples

Unweighted Unifrac Distance metric

The two dimensional plot of Principal Coordinate analysis performed using the unweighted Unifrac distance metric (Figure 8A). The plot is clustered by the outcome variable i.e., either Positive, Negative, Null implantation result, as it can be seen, no significant clustering can be observed between the samples across the 3 outcome variables.

Even in the weighted PCoA, the clustering is not clearly visualized among the 3 outcome variables (Figure 8B).

Vaginal Samples

Unifrac Distance metric

A similar analysis was carried out in the vaginal samples, resulting in the PCoA plots for unweighted and weighted metric (Figure 9). Here, some level of clustering can be observed from the 2-Dimensional PCoA plot in PC1vsPC2 (Figure 9A) and PC1vsPC3 plot, indicating distinct differences in the communities. The

positive and negative samples are separable with a few exceptions. No significant clustering is observed for the three different outcome variables (Figure 9B).

The distance comparison using boxplot is also informative as boxplot depicts the comparison within the different levels or categories of the metadata variables. Within samples distances when compared across the 3 categories, are not very different (Figure 10). However, the positive and negative samples have similar ranges of distances and the null samples have a very small range. In the weighted UniFrac boxplot, it can be observed that though the distances are very similar indicating that the three groups are nearly identical, the range of distances is also based on the relative abundance of the community (Figure 10Aa and Ba).

Irrespective of whether the abundances are taken into consideration (weighted) or not (unweighted), the 3 communities are very similar to each other in terms of the UniFrac distance. The Positive and Null groups show significantly different means. In the unweighted boxplot, the distances seem to be higher compared to that of weighted boxplots (Figure 10Ab and Bb). The range of the distances, however, seem to be higher in the weighted box plot indicating large within sample variations due to consideration of abundance in the calculation of the distance.

Conclusion

Genera specific to implantation failure resulting from this study was validated using previous literature of its pathogenicity. These genera were mainly *Prevotella, Proteus, Staphylococcus, Lactococcus, Planococcaceae and Haemophilus*. Further research on the mechanism of action of these genera to bring about pathogenic infections as well as elevated levels of risk for implantation failure in assisted reproductive techniques will further strengthen these findings. Additionally, numerous pathogenic microbial floras were also observed in the cases leading to successful implantation. Interestingly, microbes of unique origin, namely from marine, soil, salt rich origin were also reported. (*Defluviitoga, Geodermatophilus, Marininema, Neorhizobium, Parcubacteria, Haererehalobacter, Pelagibacteraceae, Rhodobacteraceae)* Possible sources from where these bacteria could have entered the cervical or vaginal site must also be evaluated. The environment rich in *Lactobacillus spp*. have been previously linked to successful implantation results. The dysbiosis in cervical and vaginal environment (generally with reduction in Lactobacillus) has been attributed to higher risk of infections, pre-term birth, miscarriage, infertility, etc.

From this study, the possible sources of contamination were evaluated to be the catheter tip during embryo transfer, the DNA extraction kits prior to sequencing. Further studies on the pathogen mechanism as well as the possible sources of contamination need to be evaluated to improve success ratio of In-vitro fertilization. This study proves to be a strong foundation for the development of certain interventions in the IVF procedure such as administration of probiotics in females undergoing the procedure to prepare a suitable environment (effectively elevate levels of *Lactobacilli*) for implantation. Another effective measure would be to first profile the metagenome of the female, and effectively target the pathogenic

microorganisms by means of antibiotics or at-site injections to diminish the unwanted bacterial growth. Several other strategies which have been proposed through several studies are thorough technique for selection of sperms, pre- implantation genetic screening, monitoring of the embryo through time lapse, etc. The functional mechanism by which the prevalent microbiota exerts its presence and disrupts the underlying common environment to bring about development of potentially pathogenic infections that lead to reduced fertility rates must be evaluated in exhaustive detail. The effective implementation of these strategies to suitable patients may result in higher levels of implantation, leading to a successful pregnancy.

List Of Abbereviation

IVF – In-Vitro Fertilization

RIF – Recurrent Implantation Failure

INSDC - International Sequence Database Collaboration

OTU – Operational Taxonomic Unit

RDP – Ribosomal Database Project

Declarations

Consent of Publication

All the authors have reviewed the manuscript and accepted to publish in the current state

Ethics approval

Ethical approval was granted from the Institutional Ethical Committee of Jilla Hospital and Research Institute, Aurangabad (approval No. JHRC/IEC-2018/1(02) dated 30/07/2018).

Conflict of Interest

Authors declare no conflict of interest

AUTHOR CONTRIBUTION

Gulab Khedkar[1§], Manju Jilla[2], CD Khedkar[4] and Pramod Yeole[10] -Concept and ideation

Gulab Khedkar[1§] - Made resources available like kits and reagents and laboratory supplies

Vidya Niranjan[3] K Piyusha[3], C Lavanya[3], Bioinformatic analysis and drafting manuscript

Manju Jilla[2], Amol Kalyankar[5], Rajshree Deolalikar[6], Chaitali Khedkar[7], Bharathi Prakash[8], P Josthna[9] - Sampling and sequencing

All the authors were involved in the reviewing of manuscript.

Acknowledgements

Authors are thankful to all staff members and students at Paul Hebert Centre for DNA Barcoding and Biodiversity Studies, Aurangabad for their assistance in completing this work.

The work was supported by Rashtriya Uchchatar Shiksha Abhiyan (RUSA), Maharashtra, India (Grant no. PD/RUSA/Order/2018/127 dated February 14th, 2018) to Prof. Gulab Khedkar is greatly acknowledged.

Funding

The work was supported by Rashtriya Uchchatar Shiksha Abhiyan (RUSA), Maharashtra, India (Grant no. PD/RUSA/Order/2018/127 dated February 14th, 2018) to Gulab Khedkar is greatly acknowledged.

DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are available in the NCBI SRA Database https://www.ncbi.nlm.nih.gov/sra/PRJNA747685

References

1. de Oliveira[1], M. S., das Chagas Medeiros, F., & Junior, J. E. (2019). Infertility and vaginal microbiome: review study. *DST-J bras Doenças Sex Transm*, *31*(1), 24-29.
2. Infertility-National Health Portal f India, 5th August 2016, https://www.nhp.gov.in/disease/reproductive-system/infertility
3. Deyhoul, N., Mohamaddoost, T., & Hosseini, M. (2017). Infertility-related risk factors: a systematic review. *Int J Womens Health Reprod Sci*, *5*(1), 24–29.
4. F., Chen, C., Wei, W., Wang, Z., Dai, J., Hao, L., … & Tang,H. (2018). The metagenome of the female upper reproductive tract. *GigaScience*, *7*(10), giy107.
5. Moreno, I., & Simon, C. (2019). Deciphering the effect of reproductive tract microbiota on human reproduction. *Reproductive medicine and biology*, *18*(1), 40–50.
6. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. – FASTQC Paper J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., … & Tiedje, J. M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic acids research*, *42*(D1), D633-D642.
7. Balvočiūtė, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT— how do these taxonomies compare?. *BMC genomics*, *18*(2), 114.

8. USEARCH tool webpage: https://www.drive5.com/usearch/ Phyloseq Documentation: https://joey711.github.io/phyloseq/ & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, *72*(7), 5069-5072
9. QIIME tool webpage: http://qiime.org/ Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963.
10. M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., … Beiko, R. G. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*, *31*(9), 814.
11. Shannon Index: https://slideplayer.com/slide/9211622/ Thukral, A. K. (2017). A review on measurement of Alpha diversity in biology. *Agric. Res.J*, *54*, 1–10.
12. RarefactionAnalysis: https://grunwaldlab.github.io/metacoder_documentation/workshop–06–quality_control.html
13. B. R., Shin, J., Guevarra, R., Lee, J. H., Kim, D. W., Seol, K. H., … & Isaacson, R. E. (2017). Deciphering diversity indices for a better understanding of microbial communities. *J. Microbiol. Biotechnol*, *27*(12), 2089-2093.
14. Gotelli, N. J., & Chao, A. (2013). Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. Beta Diversity; Dan Knights
15. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *The ISME journal*, *5*(2), 169–172.
16. Datta, A. K., Campbell, S., Deval, B., & Nargund, G. (2015). Addons in IVF programme–Hype or Hope?. *Facts, views & vision in ObGyn*, *7*(4), 241
17. Almazán, J., … & Ramon, D. (2016). Evidence that the endometrial microbiota has an effect on implantation success or failure. *American journal of obstetrics and gynecology*, *215*(6), 684-703.
18. Lai, Q., Wang, J., Gu, L., Zheng, T., & Shao, Z. (2013). Alcanivorax marinus sp. nov., isolated from deep-sea water. *International journal of systematic and evolutionary microbiology*, *63*(12), 4428–4432.
19. Pathak, P., Trilligan, C., & Rapose, A. (2014). Bifidobacterium—friend or foe? A case of urinary tract infection with Bifidobacterium species. *Case Reports*, *2014*, bcr2014205122
20. Bhaskar, M. M., Sistla, S., & Kumaravel, S. (2017). A case of pyometrocolpos with Bifidobacterium species. *Anaerobe*, *44*, 48–50.
21. Shukla, S. K., Harney, M., Jhaveri, B., Andrews, K., & Reed, K. D. (2003). Is a black- pigmented Corynebacterium species an opportunistic pathogen during pregnancy? Literature review and report of 3 new cases. *Clinical infectious diseases*, *37*(6), 834–837.
22. J., Steinsland, H., Tumwine, J. K., Namugga, O., Mukunya, D., Bwanga, F., … Nankabirwa, V. (2020). Vaginal colonisation of women in labour with potentially pathogenic bacteria: a cross sectional study at three primary health care facilities in Central Uganda. *BMC Infectious Diseases*, *20*(1), 98.

23. Hania, W. B., Godbane, R., Postec, A., Hamdi, M., Ollivier, B., & Fardeau, M. L. (2012). Defluviitoga tunisiensis gen. nov., sp. nov., a thermophilic bacterium isolated from a mesothermic and anaerobic whey digester. *International journal of systematic and evolutionary microbiology*, *62*(6), 1377–1382. differential distribution of bacteria between cancerous and noncancerous ovarian tissues in situ. *Journal of Ovarian Research*, *13*(1), 8.
24. Ruggeri, M., Cannas, S., Cubeddu, M., Molicotti, P., Piras, G. L., Dessole, S., & Zanetti,S. (2016). Bacterial agents as a cause of infertility in humans. *New Microbiol*, *39*(3), 206–209. (2018). Impact of asymptomatic genital tract infections on in vitro fertilization (IVF) outcome. *PloS one*, *13*(11).
25. Kodati, V. L., Govindan, S., Movva, S., Ponnala, S., & Hasan, Q. (2008). Role of Shigella infection in endometriosis: a novel hypothesis. *Medical hypotheses*, *70*(2), 239–243.
26. Onderdonk, A. B., Delaney, M. L., & Fichorova, R. N. (2016). The human microbiome during bacterial vaginosis. *Clinical microbiology reviews*, *29*(2), 223–238.
27. Jin, L., Lee, H. G., Kim, H. S., Ahn, C. Y., & Oh, H. M. (2013). Geodermatophilussoli sp. nov. and Geodermatophilusterrae sp. nov., two actinobacteria isolated from grass soil. *International journal of systematic and evolutionary microbiology*, *63*(7), 2625–2629.
28. Moreno, I., & Simon, C. (2018). Relevance of assessing the uterine microbiota in infertility. *Fertility and sterility*, *110*(3), 337–343.
29. García-Velasco, J. A., Menabrito, M., & Catalán, I. B. (2017). What fertility specialists should know about the vaginal microbiome: a review. *Reproductive biomedicine online*, *35*(1), 103–112.
30. Mastromarino, P. (2018). Role of lactobacilli and lactoferrin in the mucosal cervicovaginaldefense.*Frontiers in immunology*, *9*,376. (2012). Marininema mesophilum gen. nov., sp. nov., a thermoactinomycete isolated from deep sea sediment, and emended description of the family Thermoactinomycetaceae. *International journal of systematic and evolutionary microbiology*, *62*(6), 1383-1388. (2005). Owenweeksia hongkongensis gen. nov., sp. nov., a novel marine bacterium of the phylum ‗Bacteroidetes'. *International journal of systematic and evolutionary microbiology*, *55*(3), 1051-1057.
31. Wu, X., Pan, J., Li, M., Li, Y., Bartlam, M., & Wang, Y. (2019). Selective enrichment of bacterial pathogens by microplastic biofilm. *Water research*, *165*, 114979.
32. Jongman, M., Chidamba, L., & Korsten, L. (2017). Bacterial biomes and potential human pathogens in irrigation water and leafy greens from different production systems described using pyrosequencing. *Journal of applied microbiology*, *123*(4), 1043–1053.
33. Kong, Y., Liu, Z., Shang, Q., Gao, Y., Zheng, C., Deng, X., & Chen, T. (2019). Revealing the Link between Vaginal Microbiota and Outcomes of in vitro Fertilization: A Clinical Trial.
34. Darwish, A. M. (2016). Unexplained Female Infertility Alert Over Overt and Hidden Genital Infections. *Genital Infections and Infertility*, 21.
35. Ranjit, E., Raghubanshi, B. R., Maskey, S., & Parajuli, P. (2018). Prevalence of bacterial vaginosis and its association with risk factors among nonpregnant women: A hospital based study. *International journal of microbiology*, *2018*. (2019). Effect of the vaginal microbiome on the pregnancy rate in

women receiving assisted reproductive treatment. *Journal of assisted reproduction and genetics*, *36*(10), 2111-2119.

36. Moreno, I., & Franasiak, J. M. (2017). Endometrial microbiota—new player in town. *Fertility and sterility*, *108*(1), 32–39.
37. Zamora, R. B., & Vizcaíno Ledesma, A. M. (2019). Microbiome and Fertility. *Ann Clin Obstet Gynecol*, *1*(1), 1001. (2017). 12: Alterations in the placental microbiome among spontaneous preterm births. *American Journal of Obstetrics & Gynecology*, *216*(1), S9-S10.
38. Khamees, S. (2012). Characterization of vaginal discharge among women complaining of genital tract infection. *Int. J. of Pharm. & Life Sci.(IJPLS) Oct*, *3*(10).
39. Peric, A., Weiss, J., Vulliemoz, N., Baud, D., & Stojanov, M. (2019). Bacterial colonization of the female upper genital tract. *International journal of molecular sciences*, *20*(14), 3405.
40. Benner, M., Ferwerda, G., Joosten, I., & Van der Molen, R. G. (2018). How uterine microbiota might be responsible for a receptive, fertile endometrium. *Human reproduction update*, *24*(4), 393–415.
41. Tomaiuolo, R., Veneruso, I., Cariati, F., & D'Argenio, V. (2020). Microbiota and Human reproduction: The case of female infertility. *High-Throughput*, *9*(2), 12.
42. Pellati, D., Mylonakis, I., Bertoloni, G., Fiore, C., Andrisani, A., Ambrosini, G., & Armanini, D. (2008). Genital tract infections and infertility. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *140*(1), 3–11.
43. Casari, E., Ferrario, A., Morenghi, E., & Montanelli, A. (2010). Gardnerella, Trichomonas vaginalis, Candida, Chlamydia trachomatis, Mycoplasma hominis and Ureaplasma urealyticum in the genital discharge of symptomatic fertile and asymptomatic infertile women. *The new microbiologica*, *33*(1), 69.
44. Pugh, G. (2018). The Role of Microbial Diversity in Female Reproductive Health: An Analysis of the Microbiome's Influence on Pregnancy Outcomes in Postparturient Mares.
45. Ndiokwere, C., Olise, N. A., Nmewurum, V., Omoregie, R., Agbakoba, N. R., & Anukam, K. C. (2019). 16S rRNA Metagenomics of Seminal Fluids from Medical Microbiology Laboratory in a Tertiary Hospital, Southern Nigeria. *Journal of Medical Laboratory Science*, *29*(3), 86–109.
46. Birdilla Selva Donio, M., Chelladurai Karthikeyan, S., Michaelbabu, M., Uma, G., Raja Jeya Sekar, R., & Citarasu, T. (2018). Haererehalobacter sp. JS1, a bioemulsifier producing halophilic bacterium isolated from Indian solar salt works. *Journal of basic microbiology*, *58*(7), 597–608.
47. Kutschera, U., & Schauer, S. (2012). Prokaryotic biodiversity in marine versus terrestrial ecosystems: Methylobacteria and research ethics. *J Marine Sci Res Dev*, *2*, e113.
48. Nielsen, M. C., & Jiang, S. C. (2019). Alterations of the human skin microbiome after ocean water exposure. *Marine pollution bulletin*, *145*, 595–603.
49. Y., … & Gupta, R. (2019). Diversity of endocervical microbiota associated with genital Chlamydia trachomatis infection and infertility among women visiting obstetrics and gynecology clinics in Malaysia. *PloS one*, *14*(11).

50. Pessach, I., … Baka, S. (2019). Genital tract infection and associated factors affect the reproductive outcome in fertile females and females undergoing in vitro fertilization. *Biomedical reports*, *10*(4), 231-237.
51. Pohlner, M., Dlugosch, L., Wemheuer, B., Mills, H., Engelen, B., & Reese, B. K. (2019). The majority of active Rhodobacteraceae in marine sediments belong to uncultured genera: a molecular approach to link their distribution to environmental conditions. *Frontiers in microbiology*, *10*, 659.
52. Drago, L., De Grandi, R., Altomare, G., Pigatto, P., Rossi, O., & Toscano, M. (2016). Skin microbiota of first cousins affected by psoriasis and atopic dermatitis. *Clinical and Molecular Allergy*, *14*(1), 2.
53. Vallès, Y., Gosalbes, M. J., de Vries, L. E., Abellán, J. J., & Francino, M. P. (2012). Metagenomics and development of the gut microbiota in infants. *Clinical Microbiology and Infection*, *18*, 21–26.
54. Nami, Y., Haghshenas, B., & Khosroushahi, A. Y. (2018). Molecular Identification and Probiotic Potential Characterization of Lactic Acid Bacteria Isolated from Human Vaginal Microbiota. *Advanced pharmaceutical bulletin*, *8*(4), 683.
55. Rodrigues, M. X., Lima, S. F., Higgins, C. H., Canniatti-Brazaca, S. G., & Bicalho, R.
56. C. (2016). The Lactococcus genus as a potential emerging mastitis pathogen group: A report on an outbreak investigation. *Journal of dairy science*, *99*(12), 9864–9874.
57. Sirota, I., Zarek, S. M., & Segars, J. H. (2014, January). Potential influence of the microbiome on infertility and assisted reproductive technology. In *Seminars in reproductive medicine* (Vol. 32, No. 01, pp. 035–042). Thieme Medical Publishers.
58. Li, J. P., Hua, C. Z., Sun, L. Y., Wang, H. J., Chen, Z. M., & Shang, S. Q. (2017). Epidemiological features and antibiotic resistance patterns of Haemophilus influenzae originating from respiratory tract and vaginal specimens in pediatric patients. *Journal of pediatric and adolescent gynecology*, *30*(6), 626–631.
59. Motamedifar, M., Malekzadegan, Y., Namdari, P., Dehghani, B., Jahromi, B. N., & Sarvari, J. (2018). The prevalence of bacteriospermia in infertile men and association with semen quality in Southwestern Iran. *Infectious disorders drug targets*. & Lewis, D. A. (2018). Endocervical and vaginal microbiota in South African adolescents with asymptomatic Chlamydia trachomatis infection. *Scientific reports*, *8*(1), 1-9.
60. Soutar, C. D., & Stavrinides, J. (2019). Molecular validation of clinical Pantoea isolates identified by MALDI-TOF. *PloS one*, *14*(11).
61. Baud, D., Pattaroni, C., Vulliemoz, N., Castella, V., Marsland, B., & Stojanov, M. (2019). Sperm microbiota and its impact on semen parameters. *Frontiers in microbiology*, *10*, 234.
62. Hu, J., Gao, Y., Zheng, Y., & Shang, X. (2018). Kf-finder: Identification of key factors from host-microbial networks in cervical cancer. *BMC systems biology*, *12*(4), 54.

Tables

Table 1. List of genera obtained

Sr. No.	Genera	Sr. No.	Genera
1.	*Alcanivorax*	17.	*Streptophyta*
2.	*Bifidobacterium*	18.	*Enterobacteriaceae*
3.	*Corynebacterium*	19.	*Rhizobiaceae*
4.	*Defluviitoga*	20.	*Haererehalobacter*
5.	*Enterococcus*	21.	*Pelagibacteraceae*
6.	*Escherichia/Shigella*	22.	*Ralstonia*
7.	*Gardenella*	23.	*Ureaplasma*
8.	*Geodermatophilus*	24.	*Rhodobacteraceae*
9.	*Lactobacillus*	25.	*Lactococcus* **(Unique to negative outcome of IVF)***
10.	*Marininema*	26.	*Haemophilus* **(Unique to negative outcome of IVF)***
11.	*Neorhizobium*	27.	*Elizabethkingia*
12.	*Owenweeksia*	28.	*Pantoea*
13.	*Parcubacteria*	29.	*Planococcaceae*
14.	*Prevotella* **(Unique to Negative outcome of IVF)***	30.	*Acinetobacter*
15.	*Proteus* **(Uni**que to negative outcome of IVF)****	31.	*Streptococcus*
16.	*Staphylococcus* **(Unique to negative outcome of IVF)***		

**** The genera reported to be pathogenic and specific to negative implantation***

Certain pathogenic genera were found in positive implantation as well

Table 2. Pathways in Cervical samples

Sr. No.	Pathways	Sub-pathways	Genes involved
1.	Environmental Information Processing; Membrane Transport; Transporters	1. ABC Transporters 2. Phosphotransferase system 3. Bacterial secretion system	1. cmpA -bicarbonate transport system substrate-binding protein 2. beta-glucoside PTS system EIICBA component 3. VirB1-type IV secretion system protein
2.	Genetic Information Processing; Replication and Repair; DNA repair and recombination proteins	Prokaryotic Type; Single strand break repair; Direct Repair	ogt, MGMT; methylated-DNA-[protein]-cysteine S-methyltransferase
3.	Nucleotide metabolism	Purine metabolism	PFAS- phospho ribosyl formyl glycinamidine synthase, purl
4.	Cellular processes, cell motility, bacterial motility proteins	Flagellar system, chemotaxis proteins, two-component system proteins	cheA

Figures

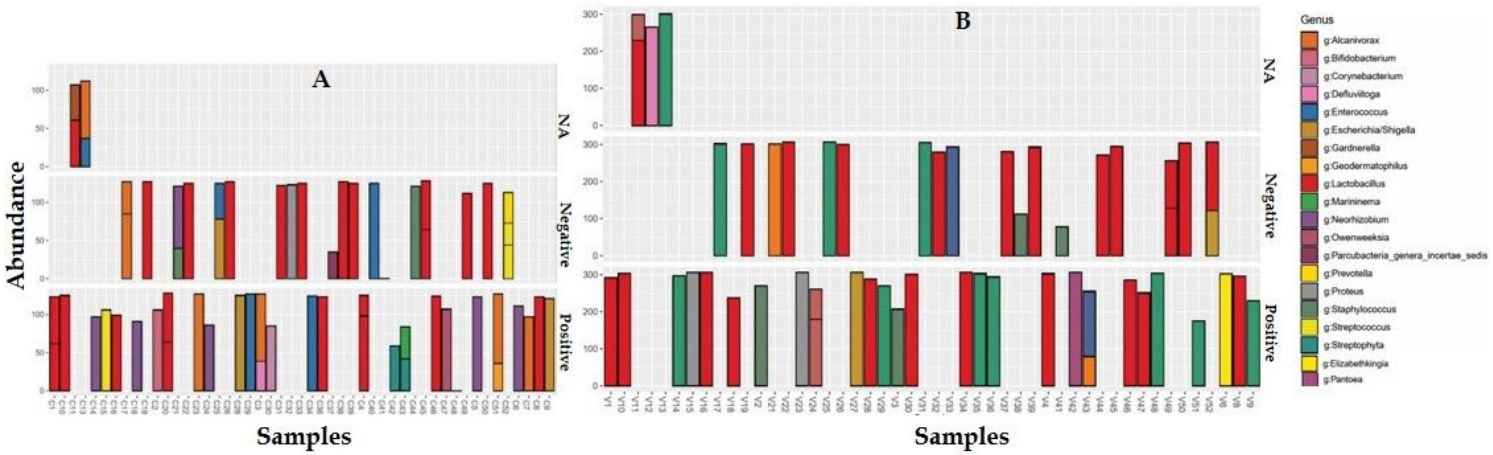

Figure 1

The bar plot for the taxonomic abundance values at genuslevel (RDP Database) A. Cervical Samples B. Vaginal Samples

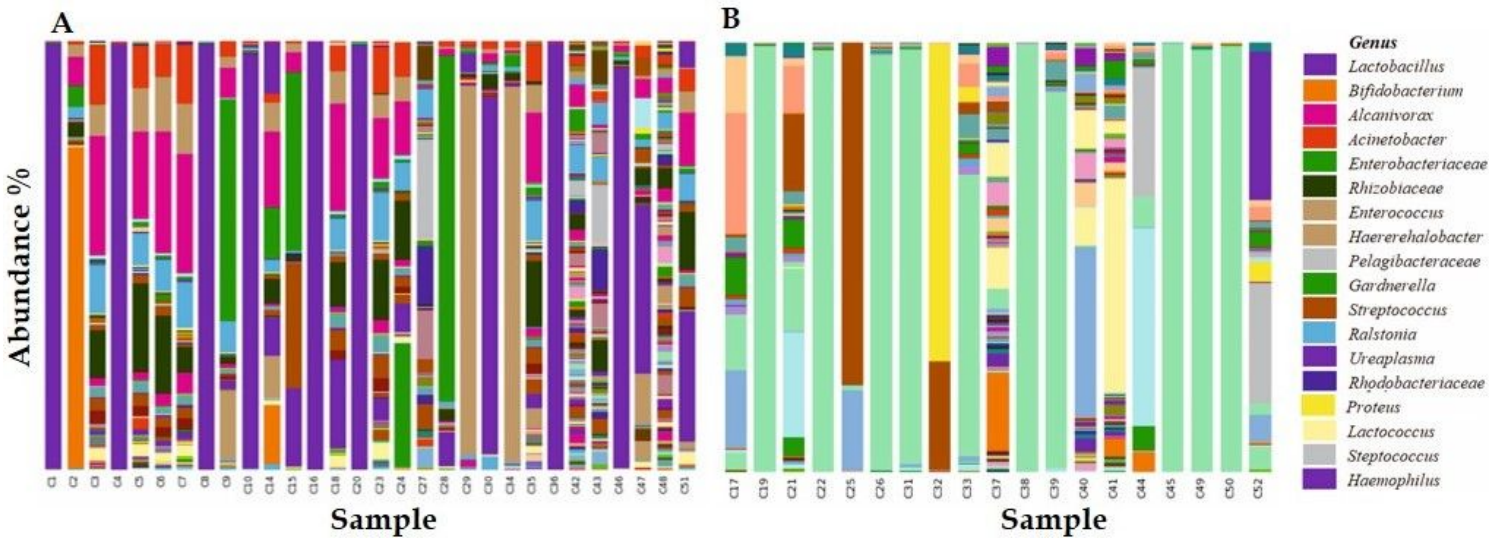

Figure 2

The bar plot for the taxonomic abundance values for the cervix samples (RDP database) A. Positive implantation B. Negative implantation

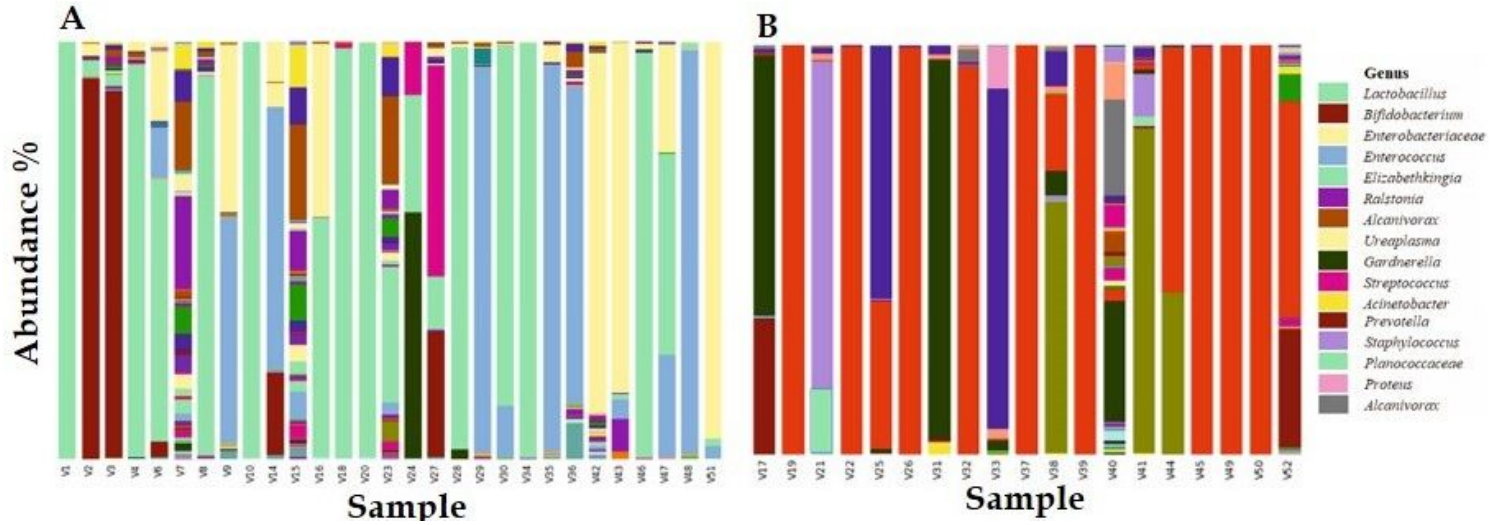

Figure 3

The barplot for the taxonomic abundance values in the vaginal samples (RDP database) A. Positive implantation B. Negative implantation

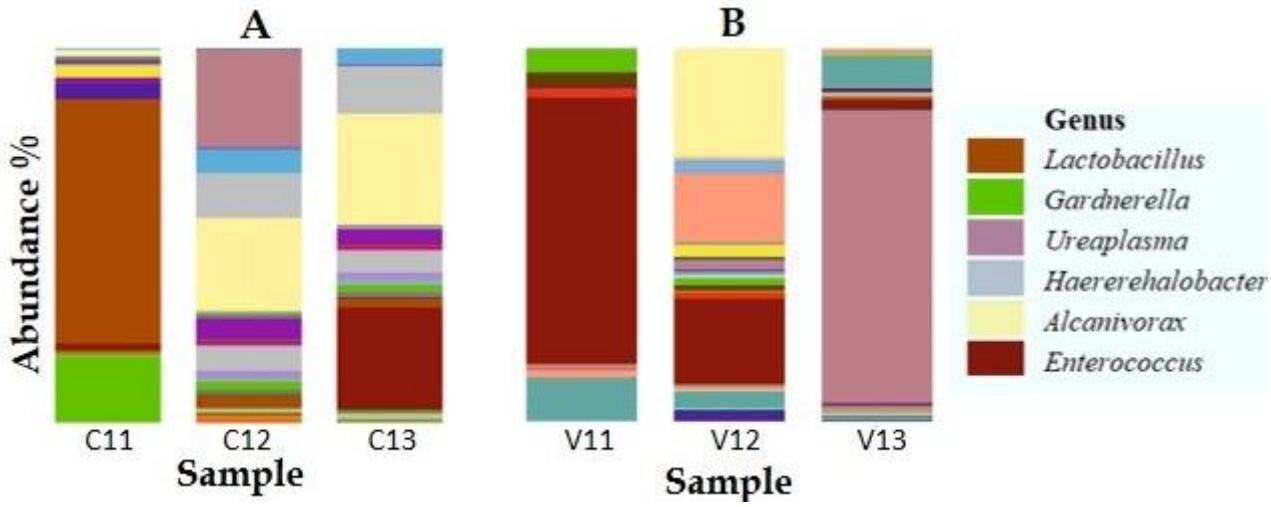

Figure 4

Microbial Communities having null impact on implantation success A. Cervix samples B. Vaginal Samples

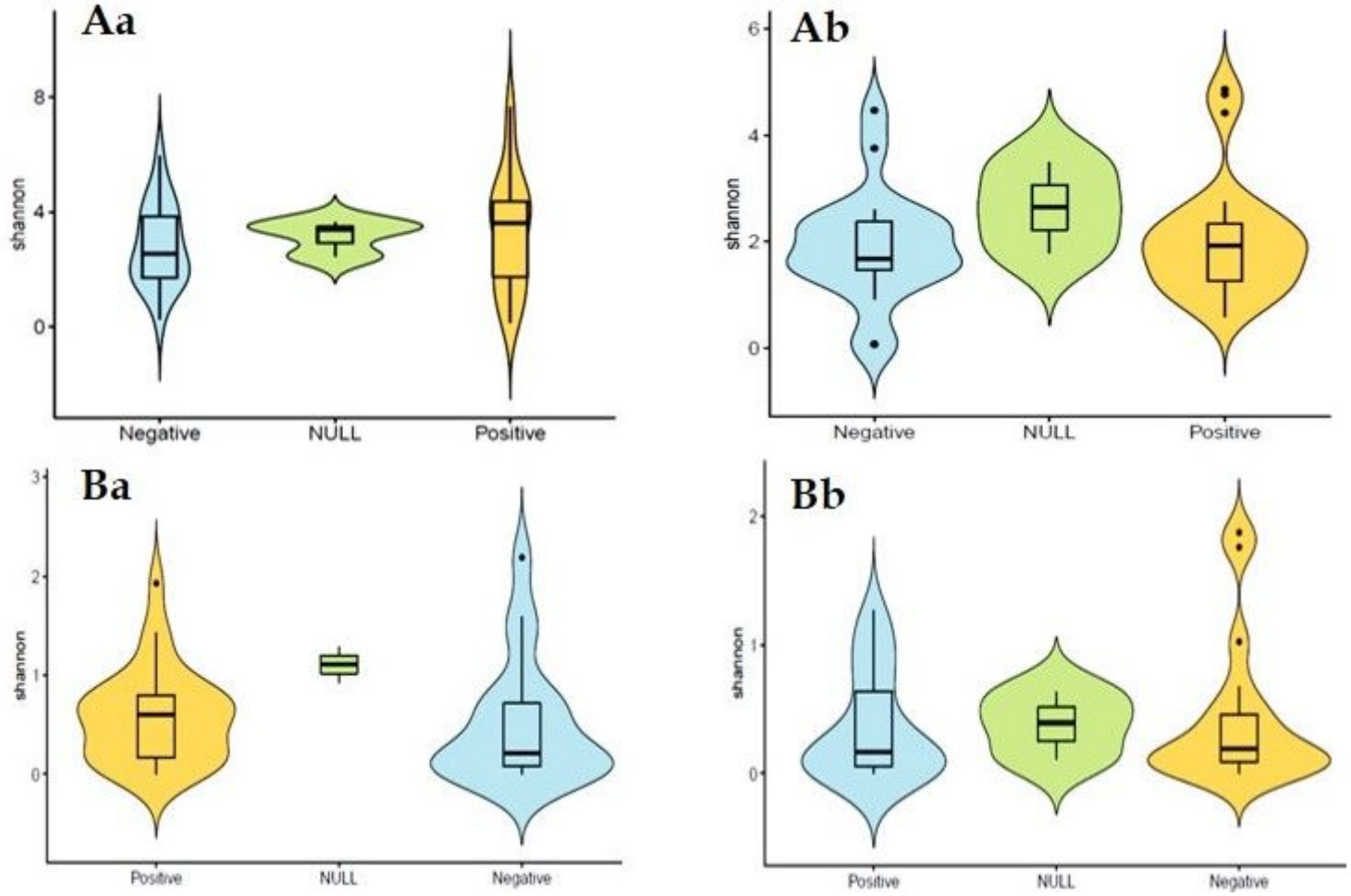

Figure 5

Alpha Diversity estimates (Shannon index) Aa. Cervix samples (Greengene database) Ab. Vaginal Samples (Greengene database) Ba. Cervix samples (RDP database) Bb. Vaginal Samples (RDP

database)

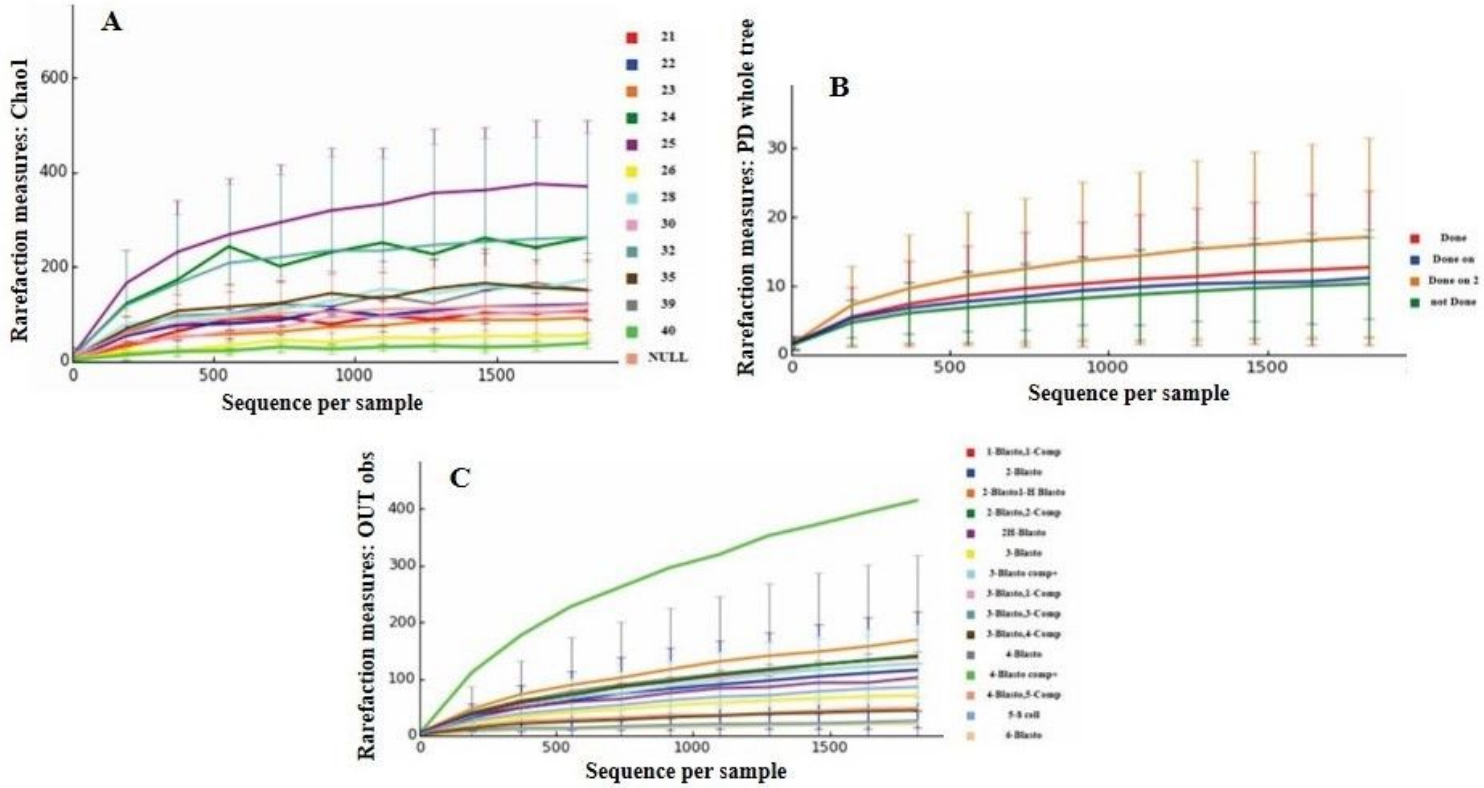

Figure 6

Rarefaction curve analysis for cervix samples A. Positive results (Chao1: age) B. Process opted for ART (PD whole tree) C. Number of embryos transferred (Observed OUT)

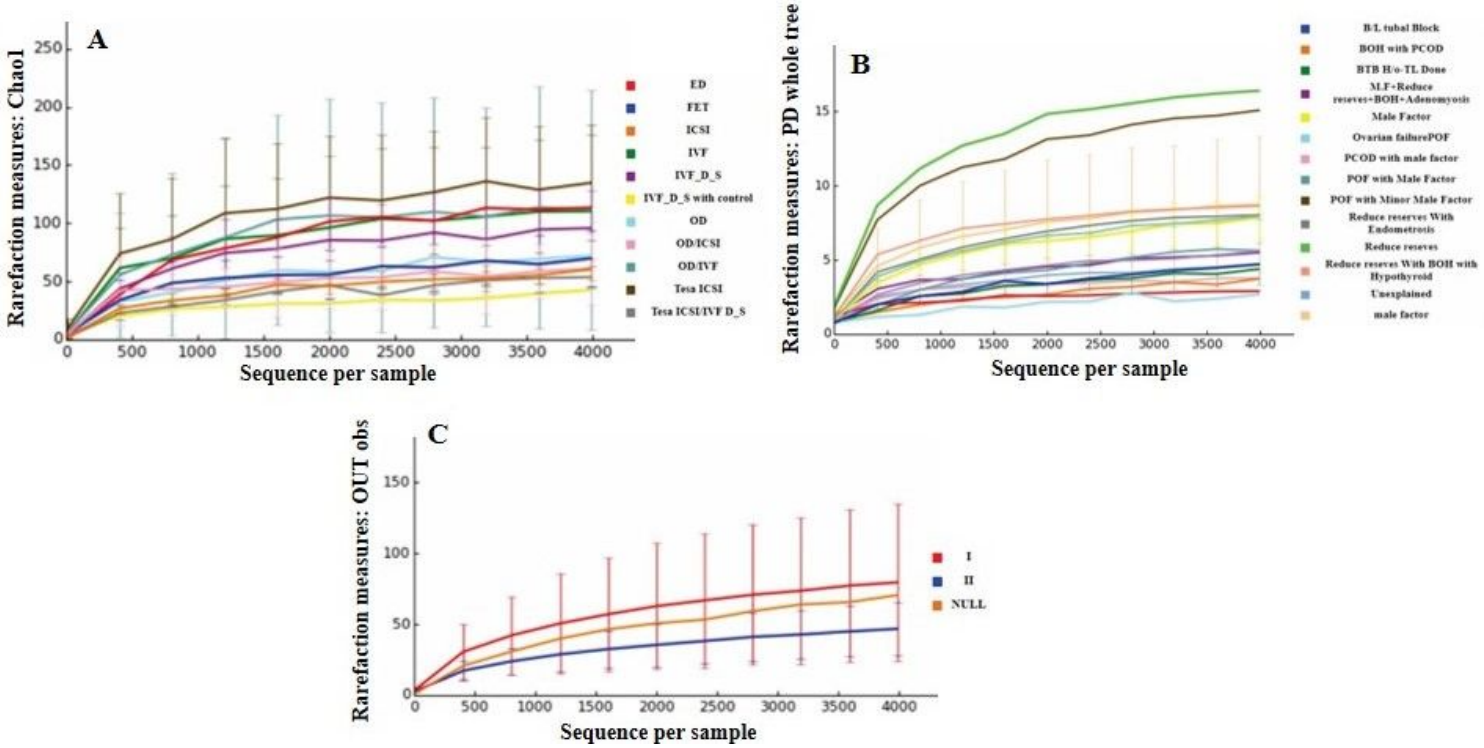

Figure 7

Rarefaction curve analysis for vaginal samples A. Procedure opted (Chao1) B. Cause of infertility (PD whole tree) C. PreSecInfertility (Observed OUT)

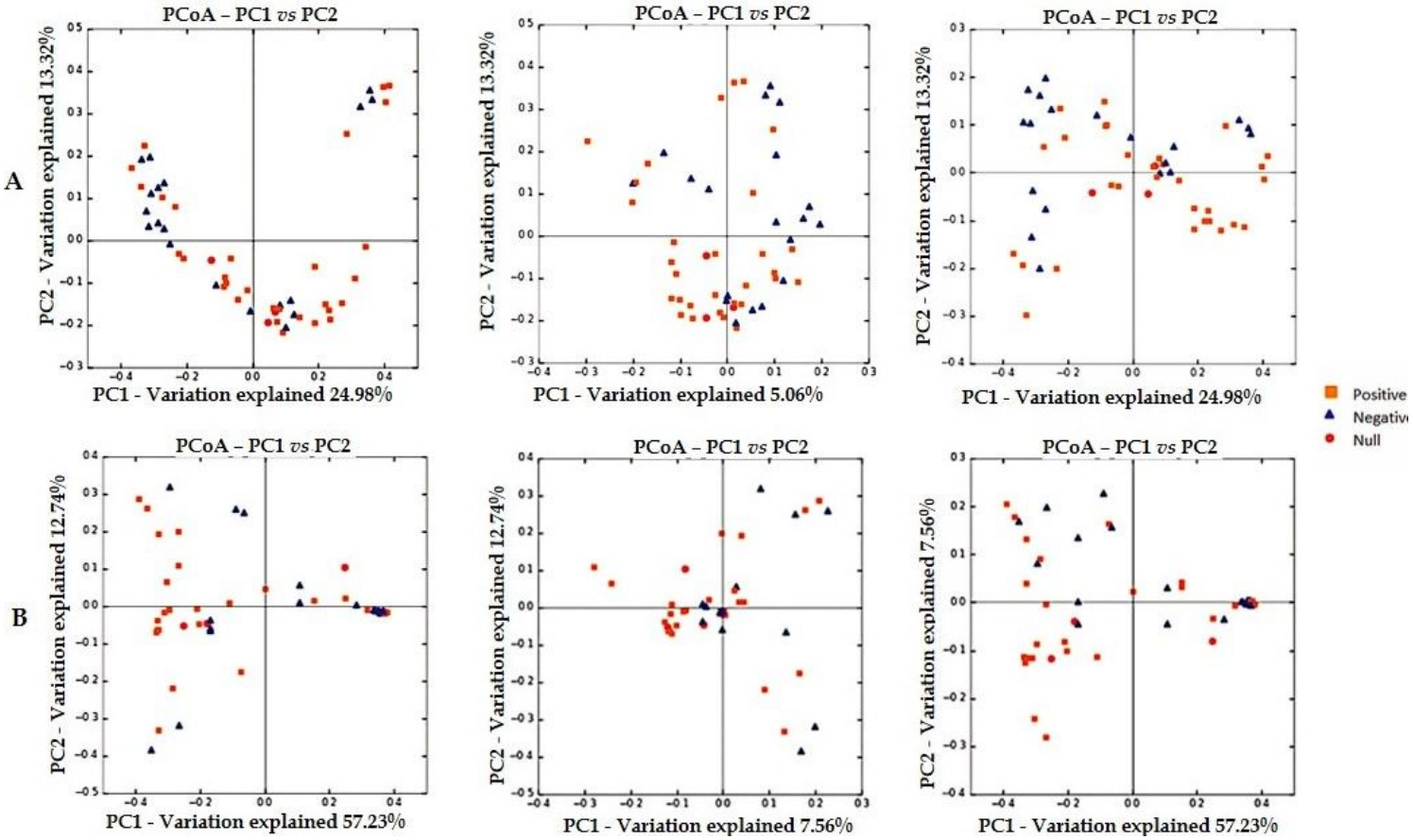

Figure 8

Two Dimensional PCoA analysis for cervix samples A. Unweighted UniFrac B. Weighted UniFrac

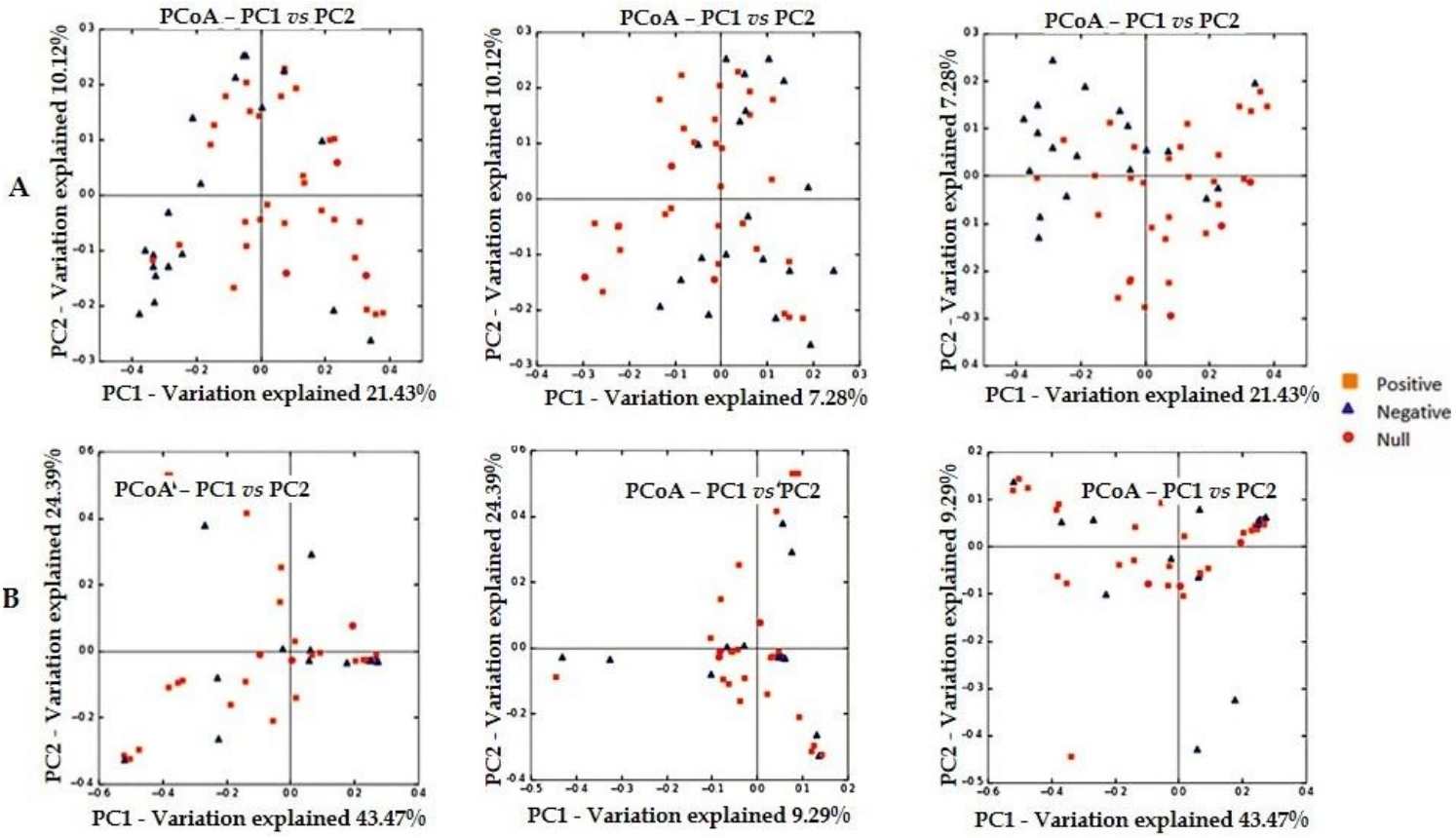

Figure 9

Two Dimensional PCoA analysis for vaginal samples A. Unweighted UniFrac B. Weighted UniFrac

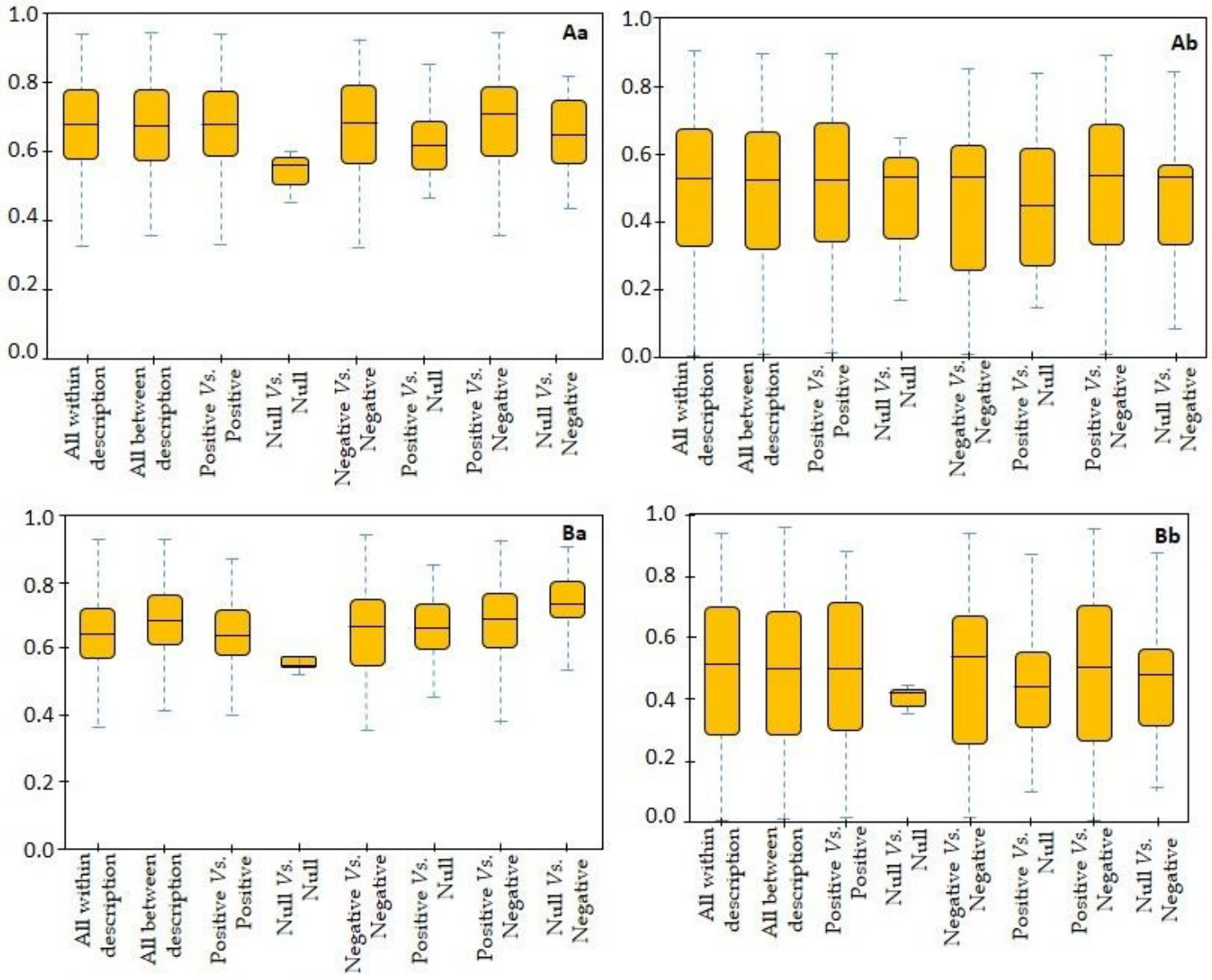

Figure 10

Distance comparison using Box plot Aa. Unweighted distance (Cervix) Ab. Weighted distance (Cervix) Ba. Unweighted distance (Vaginaal) Bb. Weighted distance (Vagina)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarytableS1.xlsx
- SupplementaryTableS2.docx
- SupplimentaryFigureS1.jpg
- SupplimentaryFigureS2.jpg

- SupplimentaryFigureS3.jpg