INTERNATIONAL ASSOCIATION OF BIOLOGICALS AND COMPUTATIONAL DIGEST
International & Peer-Reviewed Journal
E-ISSN: 2583-3995

COMPUTATIONAL META-ANALYSIS OF CERVICAL CANCER USING AVAILABLE 16S RRNA NGS DATA

[1*]Harsha Motwani, [1,2]Naman Mangukia,[1]Harshida Gadhavi, [1]Saumya K Patel and [1] Hitesh A. Solanki

[1]Department of Botany, Bioinformatics and Climate Change Impacts ManagementGujarat University, Ahmedabad, Gujarat
[2]Bioinnovations, Mumbai, India Email id : husolanki@yahoo.com

ABSTRACT

Cervical cancer is one of the most frequently occurring and deadliest gynaecologicalcancer which develops in cervical cells. Since it develops in tissues lining the internalorgans it is a Carcinoma. Human papilloma virus infection is found to top the list of carcinogenic factors. Overexpression of certain proteins due to HPV integration in host body over a time can result in carcinoma. However vaginal microbiota plays a key role in development, persistence and progression of infections leading to diseasessuch as cervical cancer. Some recent studies have revealed potential roles of microbiome in cervicovaginal diseases. Thus a comparative metagenomic studyamong such samples can uncover microbial diversities present in these samples. Dueto presence of highly conserved regions as well as hyper variable regions 16s rRNA gene sequence is selected for identification and classification of bacterial diversity. For the purpose of metagenomic analysis 16s rRNA gene sequences were analysed using QIIME pipeline.

Keywords: *16s rRNA, metagenomics, microbiota, cervical cancer.*

INTRODUCTION

Cervical cancer is one of the most frequently occurring and deadliest gynaecological cancer whichdevelops in the cells of cervix [5]. Since it develops in tissues lining the internal organs it is a Carcinoma. Normal cells undergo some changes due to some external factors or mutations. The abnormal changes occurring in cervical cells are called Squamous Intraepithelial Lesions (SIL). Cells become abnormal which can later result in cancer. These conditions are also known as cervical intraepithelial neoplasia (CIN). Depending on the extent of damage CIN can be CIN 1, CIN 2, or CIN 3 [6].

In both men and women, presence of Human papillomaviruses HPV is considered as a cause for most of the sexually transmitted diseases. Papillomaviruses are reported as a causative agent for many reproductive tract infections or STIs. Till now, on the basis of genomic differences more than 200 types of HPV have been recognised.

Type 16, 18, 31 and 45 have been reported most commonly occurring in malignant cells of cervical cancer. Type 16 accounting for about 50% of the cases. 99.7% of the cervical squamous cell carcinoma cases worldwide have shown the implication of HPV.

Metagenomics is an unbiased and reasonable method of recognising and reviewing micro-organisms within their environment and surroundings. Metagenomics has expanded the scope of targeting microbes inducing various diseases and infections including different types of cancers. The meta-genomic analyses of micro-organisms customarily aids in recognizing the microbial communities in samples and their evolutionary relationships. Genetics, environmental factors, lifestyle collectively cause different diseases. One such factor for causing different types of cancersis presence of different microbial communities in different proportions in an individual.

A healthy vaginal microbiota mainly includes Lactobacillus spp., which upholds vaginal wellbeingby dropping pH in vagina by secreting lactic acid. Bacterial Vaginosis (BV) can result in the lossof inherent Lactobacillus spp. Also promotes over-population of anaerobic bacteria which can cause inequity in vaginal microbiome allied with various health issues like vaginal

discharge syndrome, poor pregnancy outcomes, pelvic inflammatory disease, post-operative wound infections. The frequency of Bacterial vaginosis makes a person more susceptible to HPV infectionPersistent HPV infection is the cause for the development and progression of cervical cancer to ahigher grade. Study conducted among women with and without HPV infection uncovers the presence of Lactobacillus in healthy women in abundace. Some other studies reveal low frequencyof Lactobacillus spp. and notable presence of Fusobacteria spp., including Sneathia spp. in HPV positive cases [17,18].

Materials and methods

2.1 Data retrieval

Based on literature study raw sequence files were downloaded from NCBI – SRA by using key word, Cervical" (https://www.ncbi.nlm.nih.gov/sra/?term=cervical**).**

Sr. No.	Source	AccessionNo.	HPV	Region	Other Conditions
1	Homo sapiens	SRR6505152	Positive	V4	-
2	Homo sapiens	SRR6505153	Negative	V4	-
3	Homo sapiens	SRR3109486	Positive	V3 &V4	No cervicallesion
4	Homo sapiens	SRR3109483	Negative	V3 &V4	Healthy
5	Homo sapiens	SRR3109479	Positive	V3 &V4	Cervicalcancer
6	Homo sapiens	SRR3109471	Positive	V3 &V4	SIL
7	Human metagenome	SRR7225960	Positive	V3 &V4	No bacterial vaginosis
8	Human metagenome	SRR7225958	Positive	V3 &V4	Bacterial vaginosis
9	Human metagenome	SRR7225944	Negative	V3 &V4	No bacterial vaginosis
10	Human metagenome	SRR7225943	Negative	V3 &V4	Bacterial vaginosis
11	Human reproductive system metagenome	ERR1684337	-	V4 &V5	Healthy
12	Human reproductive system metagenome	SRR3506081	Positive	V3 &V4	Normal
13	Human reproductive system metagenome	SRR3504686	Positive	V3 &V4	CIN 1
	Human reproductive system			V3 &V4	

14	metagenome	SRR3504755	Positive		CIN 2
15	Human reproductive system metagenome	SRR3504954	Positive	V3 &V4	CIN 3

Table 1 Data selected from various sources

2.2 Qality Filtering

Raw files downloaded in Sra format were converted to Fastq format using Sratoolkit version 2.9.2-ubuntu 64. Paired end fastq files were split into two files by using command fastq-dump –split- files. Quality check of these files was done by using FastQC v_11.8 tool [27].

2.3 Merge paired end reads

By using FLASH_1.2.11 the paired end files were merged.

2.4 High quality reads

The files obtained after using FLASH were again processed for Quality check so as to obtain onlyhigh quality data. The poor quality reads were trimmed by using trimmomatic software [28]. In some samples few reads were removed by using a PERL Script.

2.5 QIIME installation

Qiime includes python cod and some additional packages. For Mac OS environment MacQIIME can be used. QIIME virtual box can be installed for windows and linux environments. QIIME canbe installed by using Miniconda. Command based installation is done.

2.6 Steps to run QIIME pipeline are as follows:

2.6.1 **Quality control**

It involves generating a mapping file, validate mapping file and demultiplexing of sequence. Quality is checked by FastQC.

Demultiplexxing of data is done by using _split_libraries and for FastQ files _fastq.py is used.

2.6.2 Chimera Removal

It involves identification of chimeric sequences and filtration of chimeric and non chimeric sequences. First a Usearch61 file is generated. USEARCH61 performs reference based chimera detection as well as de novo chimera detection on the basis of abundances of input sequences [24,25].

2.6.3 OTU clustering

Sequences are clustered in into operational taxonomic units (OTU).QIIME provides different strategies for OTU clustering [24,25].

2.6.4 Selecting a representative sequence **&** Assign Taxonomy

Assignment of taxonomy to OTU representative sequences is included as a default step in the open-reference OTU-picking workfl ow. QIIME uses curated database to assign taxonomy when it uses closed-reference OUT-picking approach. Behaviour of taxonomy assignment can bechanged by using parameters with the script *assign_taxonomy.py.*

2.6.5 Generate taxa Summary tables

OTU table file will be generated by _otu_table. By using this file downstream analysis is done. Inthis step out table is created, OTUs are filtered, taxonomy summary is generated and taxa plot arealso generated [25,26].

RESULTS

Source 1 HPV negative

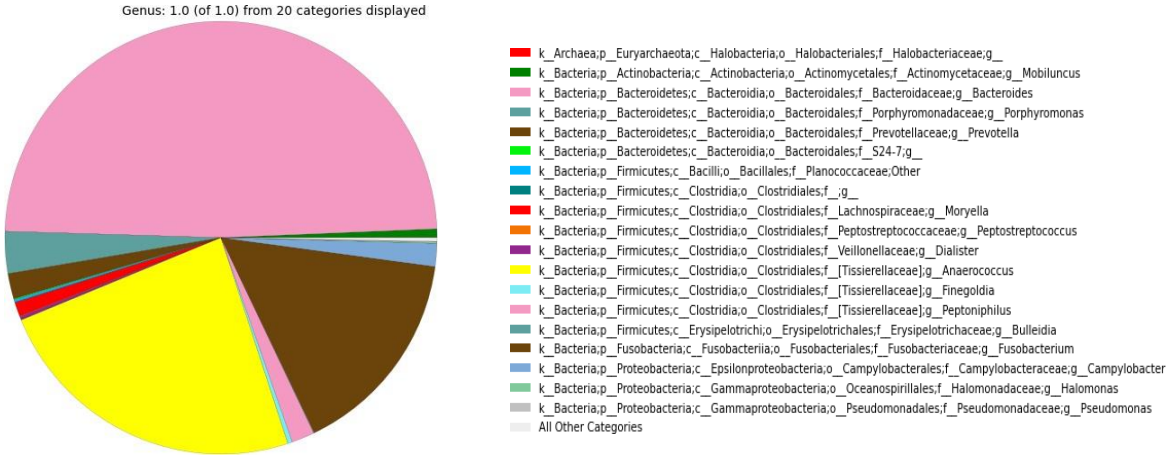


Figure 1 bacterial diversity in HPV negative sampe

Source 1 HPV positive

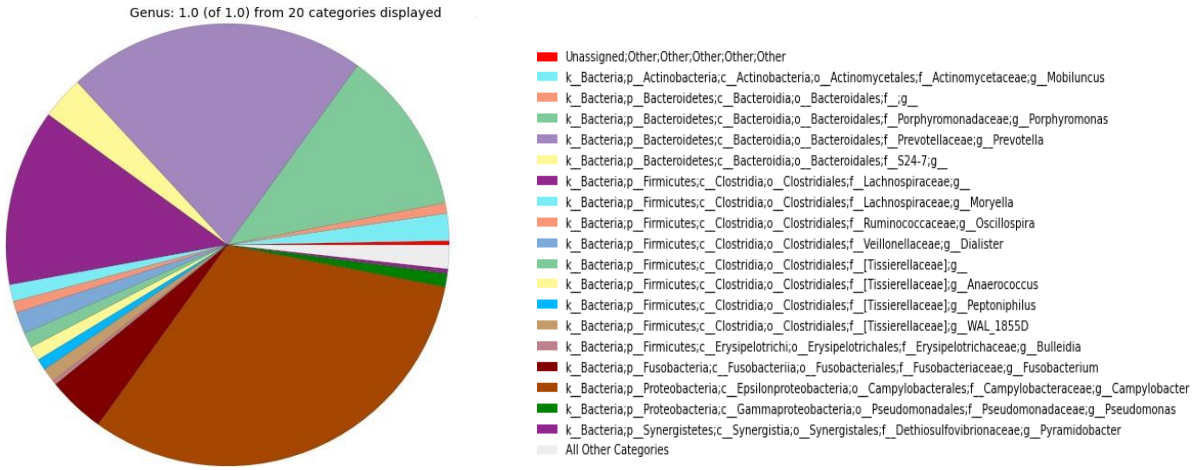


Figure 2 Bacterial diversity in HPV positive sample

Source 2 Cervical cancer

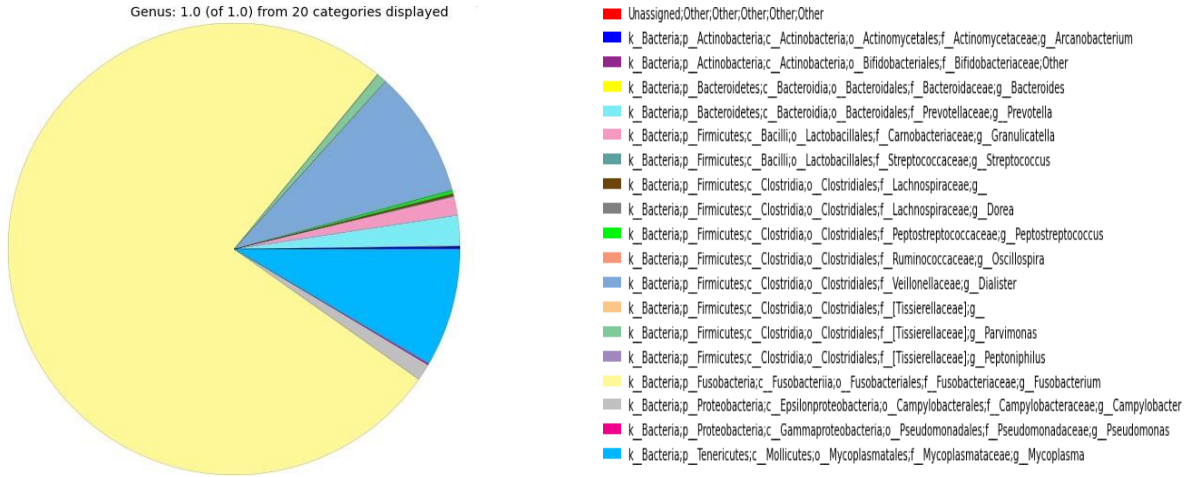


Figure 3 Bacterial diversity in cervical cancer sample

Source 2 negative healthy

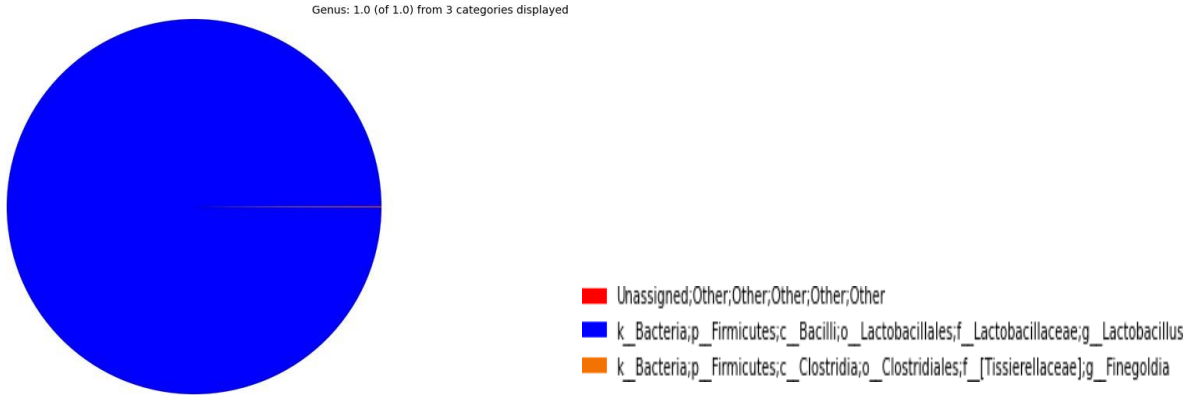
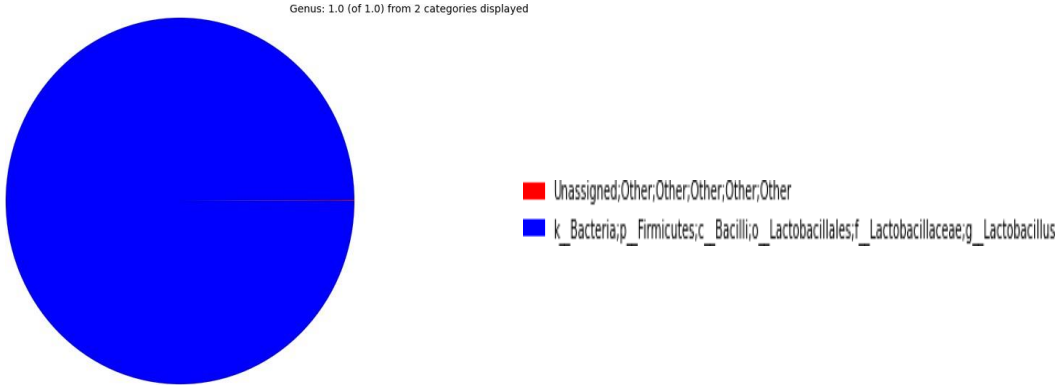


Figure 4 Bacterial diversity in HPV negative and healthy sample

Source 2 no lesion positive



Source 2 SIL

Figure 5 Bacterial diversityin HPV posituve sample without lesion

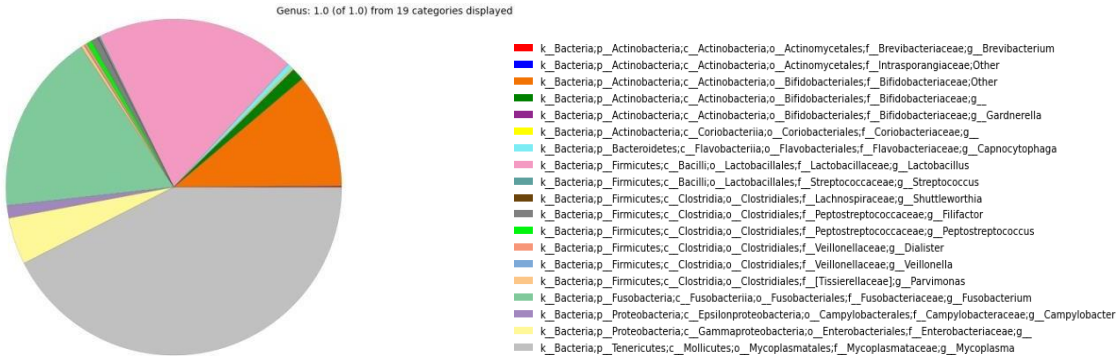


Figure 6 Bacterial diversity in SIL sample

Source 3 HPV negative BV negative

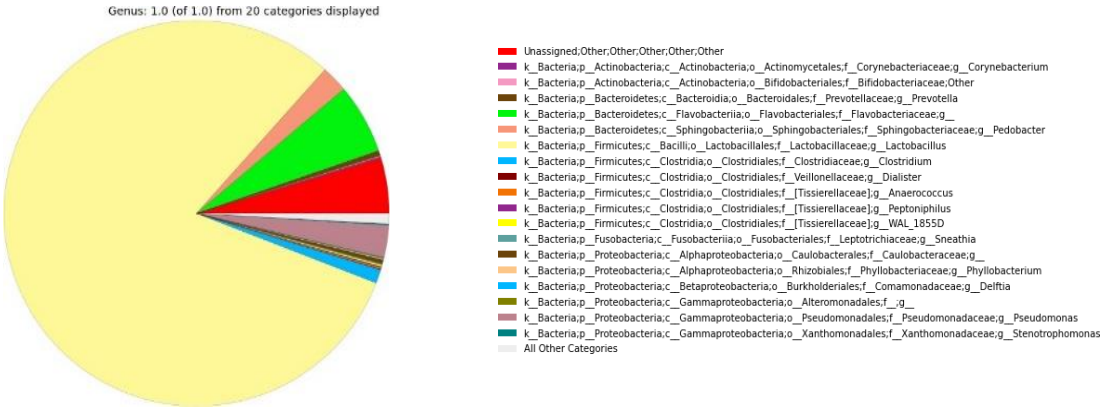


Figure 7 Bacterial diversity in HPV negative and BV negative sample

Source 3 HPV negative BV positive

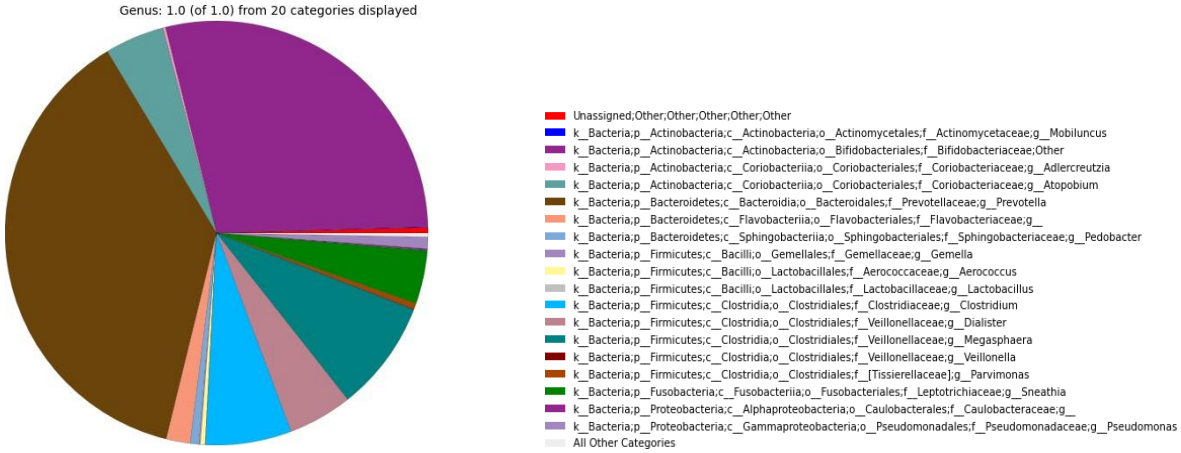


Figure 8 Bacterial diversity in HPV negative and BV positive sample

Source 3 HPV positive BV negative

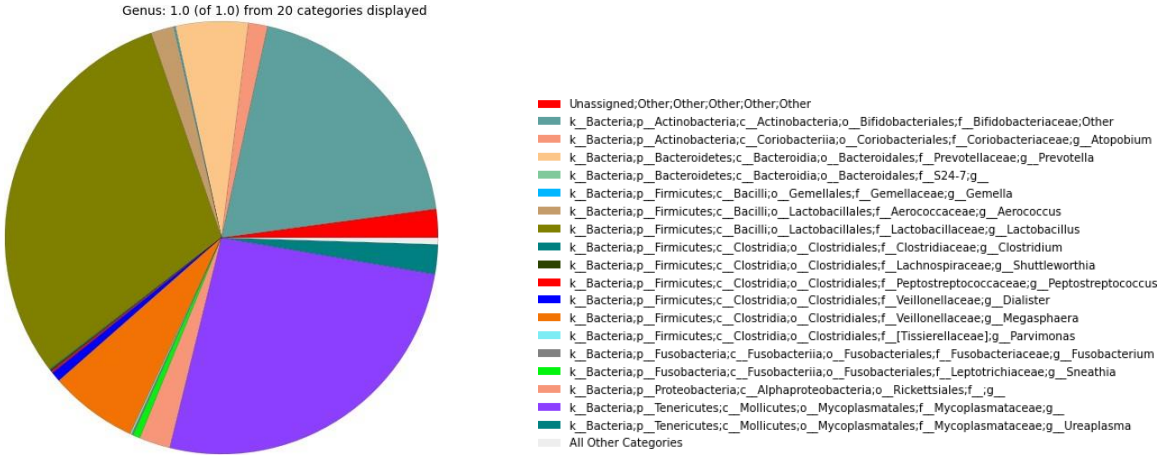


Figure 9 Bacterial diversity in HPV positive and BV negative sample

Source 3 HPV positive BV positive

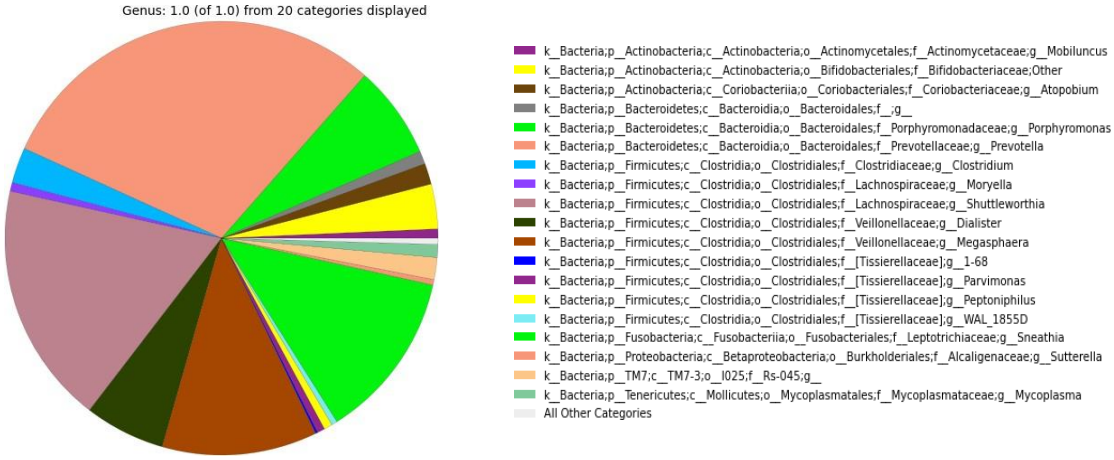
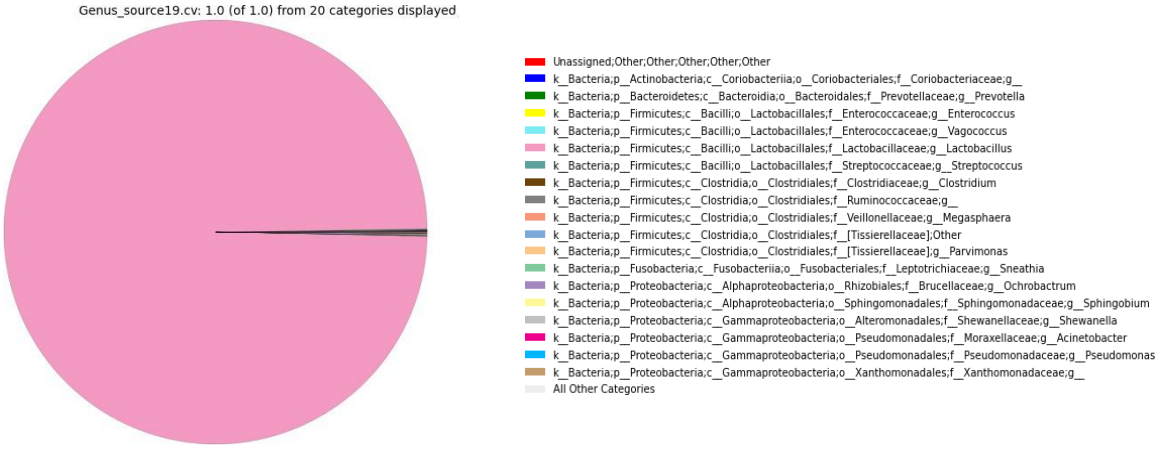


Figure 10 Bacterial diversity in HPV positive and BV positive sample

Source 4 healthy normal



Source 5 CIN 1

Figure 11 Bacterial diversity in Healthy normal sample

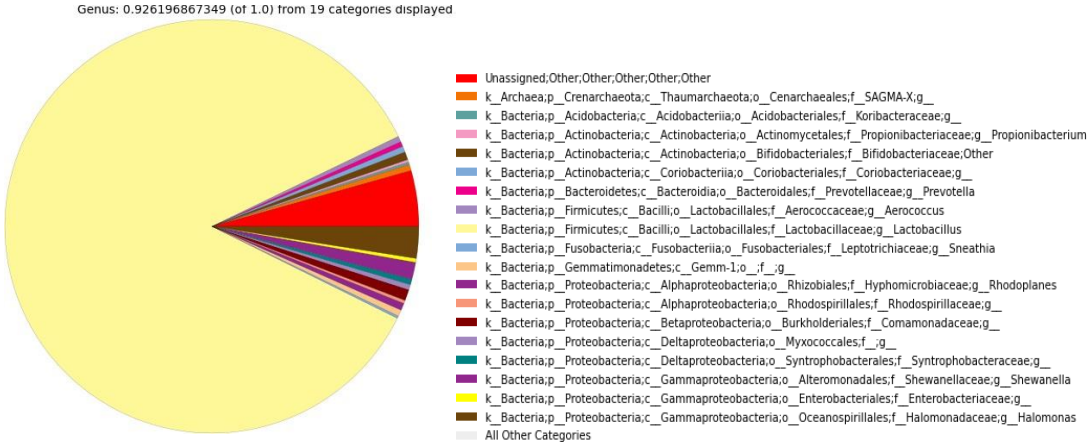
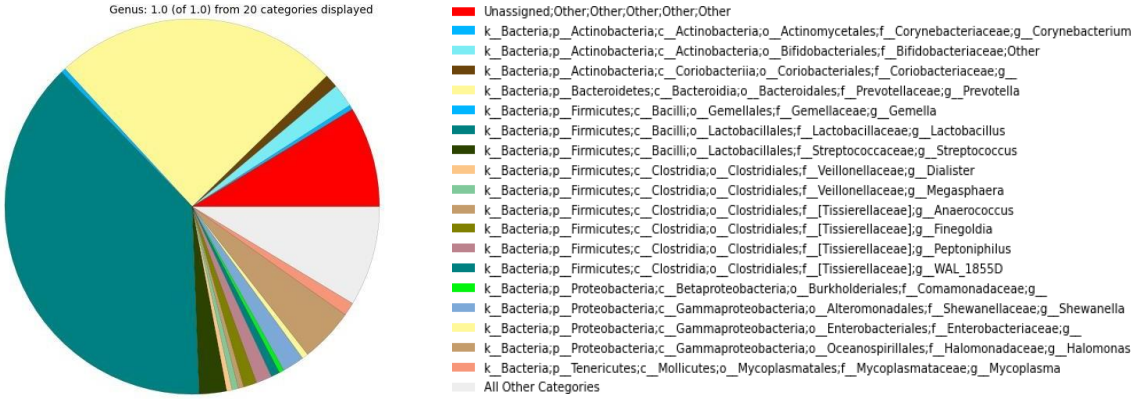


Figure 12 Bacterial diversity in CIN 1 sample

Source 5 CIN 2



Source 5 CIN3

Figure 13 Bacterial diversity in CIN 2 sample

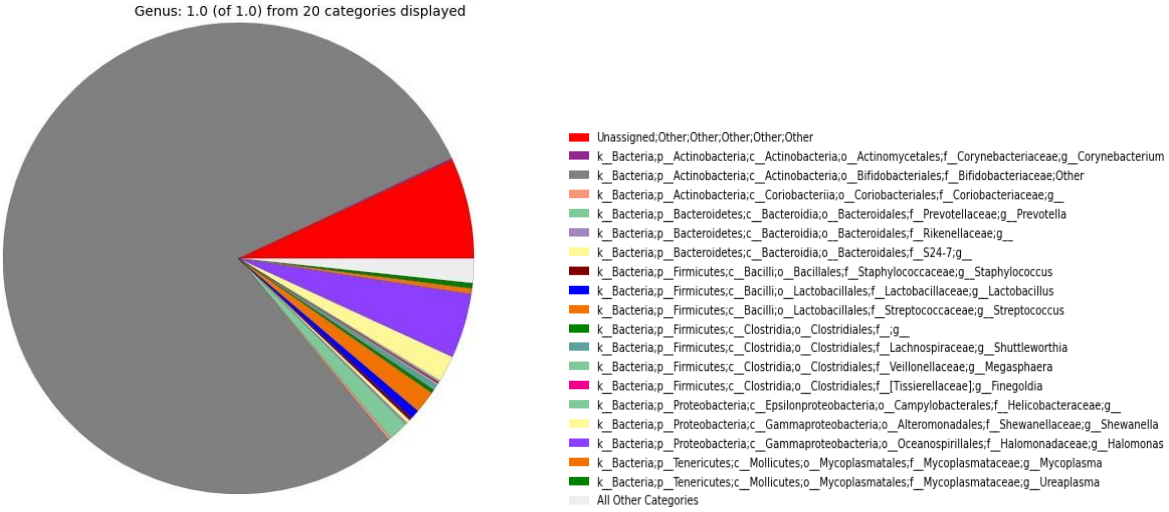
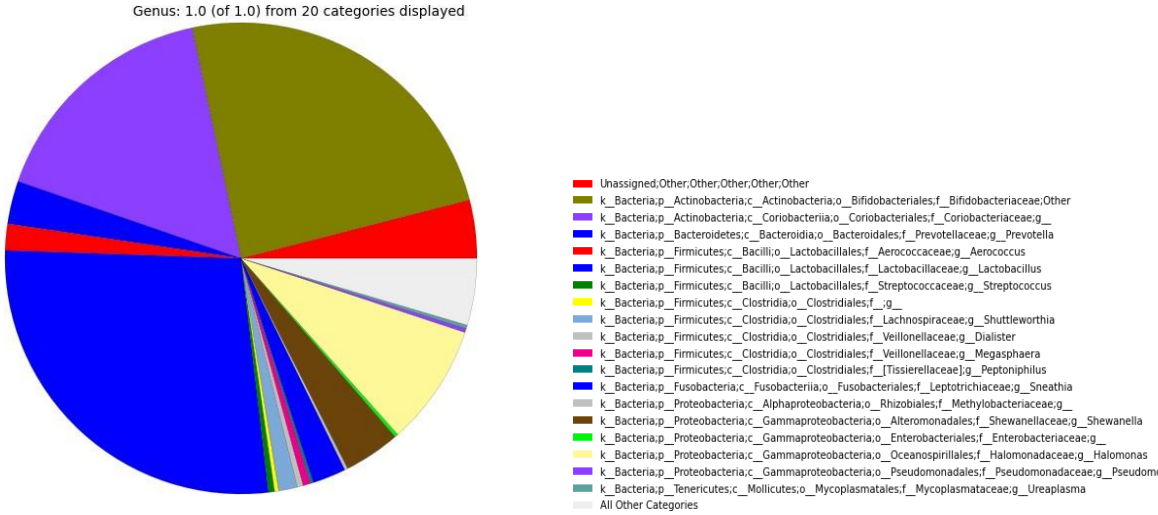


Figure 14 Bacterial diversity in CIN 3 sample

Source 5 Control



Discussion

Figure 15 Bacterial diversity in Control sample

Cervicovaginal microbiota plays a role in maintaining vaginal health. For identification of cervicalmicrobial diversities metagenomics approach was usded. This meta-analysis includes comparative studies of healthy samples, HPV positive samples, HPV negative samples, samples with differentdegrees of CIN and bacterial vaginosis samples which were downloaded from SRA. Pre- processing and quality filtering provided high quality data (QV > / = 20)which was then analysedthrough QIIME pipeline. In this study 16s rRNA gene sequences were used for identification andtaxonomical classification of microbial world occurring in healthy as well as infected samples so as to identify potential pathogens present in cervical cancer.
Lactobacillus species which is known to maintain vaginal health was found in healthy samples with relative abundance > 99%. Bacterial vaginosis samples were also having greater abundance
i.e. 80% of Lactobacillus species.

Genus	Hpv negative	HPV positive
Bacteroides	48.87%	--
Anaerococcus	23.79%	1.04%
Fusobacterium	15.76%	4.22%
Porphyromonas	3.13%	12.00%
Prevotella	1.91%	21.81%
Campylobacter	1.72%	31.87%
Lachnospiraceae	--	12.91%
Peptoniphilus	1.71%	--
Moryella	1.07%	1.23%

Table 2 Comparative results of HPV negative and HPV positive

Genus	Cervical cancer	HPV negative	HPV positive No Lesion	SIL
Fusobacterium	76.16%	--	--	17.45%
Dialister	9.10%	--	--	--
Mycoplasma	8.46%	--	--	42.52%
Prevotella	2.16%	--	--	--
Granulicatella	1.33%	--	--	--
Campylobacter	1.20%	--	--	1.20%
Lactobacillus	--	99.89%	99.93%	19.19%
Bifidobacteriaceae	--	--	--	1.14%

Table 3 comparative results of Cervical Cancer, HPV negative, HPV positive-no lesion and SIL

Genus	HPV negativeBV negative	HPV negativeBV positive	HPV positiveBV negative	HPV positiveBV positive
Lactobacillus	80.73%	--	30.13%	--
Flavobacteriaceae	5.86%	1.83%	--	--
Pseudomonas	2.68%	--	--	--
Pedobacter	2.24%	--	--	--
Clostridium	1.14%	6.57%	--	2.65%
Prevotella	--	37.54%	5.39%	29.72%
Bifidobacteriaceae	--	28.40%	19.47%	--
Megasphaera	--	8.47%	6.59%	11.59%
Dialister	--	4.91%	--	6.02%
Atopobium	--	4.55%	--	--
Sneathia	--	4.09%	--	12.59%

Shuttleworthia	--	--	--	18.04%

Table 4 Comparative results of HPV neggative- BV positive & negative, HPV positive – BV positive & negative samples

Genus	Healthy Control
Lactobacillus	99.46%
Sneathia	0.09%
Prevotella	0.08%

Table 5 Microbial diversity in Healthy control sample

Genus	CIN 1	CIN 2	CIN 3	Control
Lactobacillus	79.20%	38.31%	0.64%	27.41%
Halomonas	2.30%	4.66%	4.44%	8.34%
Shewanella	1.24%	1.91%	1.81%	3.87%
Comamonadaceae	0.81%	--	--	--
Prevotella	--	24.63%	1.37%	2.95%
Bifidobacteriaceae	--	2.00%	79.03%	24.33%
Streptococcus	--	2.36%	1.53%	--

Table 6 Comparative results of CIN 1, CIN 2, CIN 3 and Control samples

Comparative analysis of HPV positive and negative samples reveals the presence of Bacteroides and Anaerococcus exclusively in HPV negative samples in a greater abundance. Whereas Prevotella, Campylobacter and Lachnospiraceae were found abundantly in HPV positive samples.Fusobacterium was present in maximum abundance i.e. 76 % in samples with cervical cancer.

BV positive samples were having a greater abundance of prevotella which was rarely present in BV negative samples.

The abundance of lactobacillus was found to decrease with increase in CIN stage i.e lactobacilluswere higher in CIN1 (70%) and their abundance decreased in CIN 3 (0.64%).

CONCLUSION

In this study, 16s rRNA gene sequences were used from SRA for identification and taxonomic classification of cervical microbiota through the metagenomics approach. For meta-analysis of thisdata QIIME pipeline was used. Relative abundance of cervical microbes was identified in HPV positive and negative samples. Exclusive presence and greater abundance of some species such as prevotella and campylobacter was found in HPV positive samples which signifies that these couldbe potentially pathogenic and their prevelance for longer time could result in cervical carcinoma. Presence of fusobacterium in greater abundance was observed in cervical cancer samples. Whereas healthy and HPV negative samples were rich in Lactobacillus species.

REFERENCES

1) https://www.medicalnewstoday.com/articles/323648.php
2) https://www.healthline.com/health/cancer#types
3) https://www.cancerresearchuk.org/what-is-cancer/how-cancer-starts/types- of-cancer
4) https://www.medicinenet.com/cancer/article.htm
5) https://www.mayoclinic.org/diseases-conditions/cervical-cancer/symptoms-causes/syc- 20352501
6) https://www.urmc.rochester.edu/encyclopedia/content.aspx?ContentTypeID=85&ContentID=p0 0549
7) https://www.cancer.org/cancer/cervical-cancer/treating.html
8) Burd, E. M. (2003). Human papillomavirus and cervical cancer. *Clinical microbiology reviews*, *16*(1), 1-17.
9) Dueñas-González, A., Lizano, M., Candelaria, M., Cetina, L., Arce, C., & Cervera, E. (2005).Epigenetics of cervical cancer. An overview and therapeutic perspectives. *Molecular cancer*, *4*(1),38.

10) Bosch, F. X., Lorincz, A., Muñoz, N., Meijer, C. J. L. M., & Shah, K. V. (2002). The causal relation between human papillomavirus and cervical cancer. *Journal of clinical pathology*, *55*(4),244-265.
11) Franco, E. L., Duarte-Franco, E., & Ferenczy, A. (2001). Cervical cancer: epidemiology, prevention and the role of human papillomavirus infection. *Cmaj*, *164*(7), 1017-1025.
12) Finocchario-Kessler, S., Wexler, C., Maloba, M., Mabachi, N., Ndikum-Moffor, F., & Bukusi,
13) E. (2016). Cervical cancer prevention and treatment research in Africa: a systematic review froma public health perspective. *BMC women's health*, *16*(1), 29.
14) Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1-8.
15) Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum.Genet.*, *9*, 387-402.
16) MacLean, D., Jones, J. D., & Studholme, D. J. (2009). Application of next-generation'sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, *7*(4), 287.
17) Banerjee, J., Mishra, N., & Dhas, Y. (2015). Metagenomics: A new horizon in cancer research.
18) *Meta gene*, *5*, 84-89.
19) Klein, C., Gonzalez, D., Samwel, K., Kahesa, C., Mwaiselage, J., Aluthge, N., ... & Angeletti,
20) P. C. (2019). Relationship between the Cervical Microbiome, HIV Status, and PrecancerousLesions. *mBio*, *10*(1), e02785-18.
21) Kyrgiou, M., Mitra, A., & Moscicki, A. B. (2017). Does the vaginal microbiota play a role in the development of cervical cancer?. *Translational Research*, *179*, 168-182.
22) MM wALBooMER, J. A. N., Acos, M. V., MANosº, M. M., xAvıER BosCH, F., & KUMMER, J. A. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer. vworldwide. *J. pathol*, *189*, 12-19.
23) Clarridge, J. E. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical microbiology reviews*, *17*(4), 840-862.
24) Yang, B., Wang, Y., & Qian, P. Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC bioinformatics*, *17*(1), 135.
25) Fournier, P. E., Drancourt, M., Colson, P., Rolain, J. M., La Scola, B., & Raoult, D. (2013). Modern clinical microbiology: new challenges and solutions. *Nature Reviews Microbiology*, *11*(8),574.
26) Rapin, A., Pattaroni, C., Marsland, B. J., & Harris, N. L. (2017). Microbiota analysis using an illumina MiSeq platform to sequence 16S rRNA genes. *Current protocols in mouse biology*, *7*(2),100-129.
27) Kuczynski, J., Stombaugh, J., Walters, W. A., González, A., Caporaso, J. G., & Knight, R. (2011). Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Currentprotocols in bioinformatics*, *36*(1), 10-7.
28) Lawley, B., & Tannock, G. W. (2017). Analysis of 16S rRNA gene amplicon sequences using the qiime software package. In *Oral Biology* (pp. 153-163). Humana Press, New York, NY.
29) Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., & Huttley, G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, *7*(5), 335.
30) Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.
31) Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957-2963.