Establishing the Taxa with Phylogenetic Profile and *in-silico-ayurvedic* Remedy of Microbiome of Bacterial Vaginosis

S. Geethashree[1](✉), I. A. Shylesh Murthy[2], and Preenon Bagchi[1,2,3]

[1] Padmashree Institute of Management and Sciences, Bengaluru, India
geethashree175@gmail.com
[2] Vasishth Academy of Advanced Studies and Research, (Sarvasumana Association), Bengaluru, India
[3] MGM Institute of Biosciences and Technology, Aurangabad, India

Abstract. Background: Bacterial vaginosis (BV) is a vaginal disease commonly caused by bacteria which most commonly causes abnormal vaginal discharge during women's reproductive-age. BV is characterized by a shift within the vaginal flora from the dominant Lactobacillus to a polymicrobial flora and is further characterized by depletion of the resident lactic-acid-producing Lactobacillus spp. BV has been associated with serious health issues, including pelvic inflammatory disease and increased susceptibility to sexually transmitted diseases (STD) but itself is not a STD. Several genes and chromosomal regions have been found to be associated with BV in various linkage analyses, case-control studies, genome-wide association studies (GWAS), and admixture mapping studies.

Methods: In this work, using BV microbiome we perform metagenome analysis which enables us to understand how the microbiome responds to the host by studying the functional analysis of genes expressed. The technology is used to determine the order of nucleotides or targeted regions of DNA or RNA. Krona pie chart is used for visualization of the taxonomy.

Results: HUMAnN2 tool is used to determine the pathways and their abundance file of microbiome. Further, homology modelling of the BV gene receptors, Leucine Aminopeptidase 3 (LAP3) and Peptidase D (PEPD) was done and ligands from ayurvedic medicinal herbs were established.

Conclusion: The taxonomy and functional information of BV microbiome are identified.

Keywords: Bacterial vaginosis · microbiome · Lactobacillus · polymicrobial flora · GWAS · metagenomics · next-generation sequencing · Ayurvedic medicinal herbs · Operational Taxonomic Units · taxonomy · modelling · Lipinski's rule of five · docking

1 Introduction

Bacterial vaginosis (BV) is a potential reason for obstetric complications and gynaecological disorders, there's substantial interest in establishing the foremost effective

© The Author(s) 2023
R. Somashekhar et al. (Eds.): ICBDS 2022, AHSR 58, pp. 178–200, 2023.
https://doi.org/10.2991/978-94-6463-164-7_14

treatment [1, 2]. BV is the most typically reported microbiological syndrome among women of childbearing age [3]. BV is characterized by a shift within the vaginal flora from the dominant Lactobacillus to a polymicrobial flora [3, 4]. BV is characterized by depletion of the resident lactic-acid-producing Lactobacillus spp. And an overgrowth of anaerobic bacteria [5]. BV has been associated with serious health issues, including pelvic inflammatory disease, increased susceptibility to sexually transmitted diseases including HIV infection & other chronic infections [5, 6]. BV is not a sexually transmitted disease (STD) [5, 6]. Symptoms include white or grey discharge from the vagina, itchiness, burning, etc. [7, 8]. Several genes and chromosomal regions have been found to be associated with BV in various linkage analyses, case-control studies, genome-wide association studies (GWAS), and admixture mapping studies [7, 8]. Pathogenic variants in genes of high and moderate penetrance, - Leucine Aminopeptidase 3 (LAP3) & Peptidase D (PEPD) such as confer modest to a high lifetime of BV [9–11].

The arrival of next-generation sequencing (NGS) technology, genomics initially was concerned with studying genomes that were tractable from the standpoint of size and repetitive content (e.g., viruses and bacteria) and with characterization of single genes associated with disease (e.g., BV, CANCER ETC) [12]. The technology is used to determine the order of nucleotides or targeted regions of DNA or RNA. Here where raw data generation is no longer a rate-limiting factor in genome-scale studies. Galaxy is an open source for NGS data analysis [12]. The pipeline used here is metagenomics analysis which enables us to understand how the microbiome responds to the host by studying the functional analysis of genes expressed [13–15]. Further, using the technique of computer aided drug design, we have tried to establish novel ligand for BV from Ayurvedic medicinal herbs.

2 Materials and Methods

Galaxy tutorial by Hiltemann S and Batut B, 2020 Analyses of metagenomics data - The global picture is used to analyze Microbiome of Bacterial Vaginosis [16]. Lactobacillus' fasta sequences SRR011131 and SRR011132 were retrieved from SRA database. Sequences were merged Merge.files, grouped using Make.group, optimized for computation using the tools Unique.seqs, Count.seqs, Screen.seqs, Align.seqs followed by Screen.seqs again, Filter.seqs and Pre.cluster. Using Pre.cluster we performed preliminary clustering of sequences and removed undesired sequences [17]. Next, we classifed the sequences into phylotypes using a training set, which is again is provided on mothur's MiSeq SOP using Classify.seqs [17].

Next, we use the output and information of Classify.seqs to determine the abundances of the different found taxa. We classify all individual sequences to get a confidence score between 0–100% using Cluster.split tool, next, we group the sequences at 97% identity threshold using Make.shared and finally, for each cluster, we determine a consensus classification using Classify.otu tool based on the classification of the individual sequences [Operational Taxonomic Units (OTUs)] and taking their confidence scores into account. This taxonomy was visualized using Krona pie chart [18].

Next, we extract the taxonomic information using MetaPhlAn2 tool [19] which uses a database of ~1M unique clade-specific marker genes (not only the rRNA genes) identified

Table 1. Summary.seq output of fasta sequences SRR011131 and SRR011132

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	0	0	0	0	1	1
2.5%-tile:	1	20	20	0	2	23184
25%-tile:	1	21	21	0	4	231832
Median:	1	47	47	0	5	463663
75%-tile:	1	263	263	3	6	695494
97.5%-tile:		1	293	293	9	9
904141						
Maximum:	1	640	640	45	43	927324
Mean:	0.985452	134.622	134.622	1.70636	4.90572	
# of unique seqs:		786909				
total # of seqs:		927324				

from ~17,000 reference (bacterial, archeal, viral and eukaryotic) genomes. Next, using HUMAnN2 tool [20] we extract the functional information. The tool produces gene family abundance, coverage, abundance of pathways as output.

Next, using the genes present in Microbiome of Bacterial Vaginosis, their 3d structure was modeled using SWISS-MODEL [21]. Phyto-compounds were downloaded from PUBCHEM. Using, molinspiration software [22], following the principles of Lipinski's rule of five, phytocompounds were selected for docking. Further, docking was performed using patchdock [23].

3 Results and Discussion

Lactobacillus' fasta sequences SRR011131 and SRR011132 data summary as per summary.seq is given in Table 1.

Sequence Alignment of our input sequences was done with an alignment of the V4 variable region of the 16S rRNA as per mothur's MiSeq SOP from the Silva reference database. The output of Align.seq is given in Table 2.

The quality of the above alignment can be understood from the log output from the summary step given in Table 3.

The taxonomic data from the output of Classify.seqs gave the classification is given in Table 4 and its visualization in Krona and pinch is given in Figs. 1 and 2 respectively.

From the taxonomy classification we have come to know the taxa. Next we move to understanding their functional information. In this part we use Shotgun metagenomic sequencing. This process allows us to researchers to comprehensively sample all genes in our microbiome and this method enables us to detect the abundance of microbes and evaluate bacterial diversity in our microbiome.

1st we analyze SRR011131 sequence and is given as input to MetaPhlAN2 tool. The tool output is a tabular file with the community structure. This output is given as input to HUMAnN2 tool. The pathways and their abundance file (output of HUMAnN2 tool) and normalized the gene family abundances table of SRR011131 sequence (1st few lines of the output) is given in Table 5(a) and 6(a) and of SRR011132 sequence is given in Table 5(b) and 6(b) respectively.

Table 2. Align.seq output

Query Name	Query Length	Template Name	Template Length	Search Method	Search Score	Alignment Method	Pairwise Alignment Length
SRR011131.1.2	21	AJ543434.1	293	kmer	42.86	needleman	21
SRR011131.1.4	45	AJ431250.1	293	kmer	10.53	needleman	-338
SRR011131.2.2	20	Z95720.1	293	kmer	23.08	needleman	20
SRR011131.2.4	44	AF129868.1	294	kmer	13.51	needleman	-337
SRR011131.3.2	21	AY331143.1	293	kmer	35.71	needleman	21
SRR011131.3.4	44	L02888.1	291	kmer	29.73	needleman	45
SRR011131.4.2	20	AY054370.1	294	kmer	15.38	needleman	20
SRR011131.4.4	51	AF402980.1	293	kmer	11.36	needleman	2
SRR011131.5.2	25	AB154304.1	293	kmer	16.67	needleman	25
SRR011131.5.4	47	U82963.1	292	kmer	10.00	needleman	3
SRR011131.6.2	20	AF502205.1	293	kmer	30.77	needleman	20
SRR011131.6.4	47	AY500141.1	295	kmer	12.50	needleman	-342
SRR011131.7.2	20	AF013534.1	294	kmer	15.38	needleman	20
SRR011131.8.2	20	AB015542.1	293	kmer	30.77	needleman	20
SRR011131.9.2	20	AY570514.1	293	kmer	30.77	needleman	20
SRR011131.9.4	44	U81720.1	293	kmer	8.11	needleman	-336

Table 3. Summary.seq output of Align.seq.

```
         Start   End     NBases  Ambigs  Polymer NumSeqs
Minimum: 0       0       0       0       1       1
2.5%-tile:       1       1231    1       0       1
 22046
25%-tile:        3766    4049    5       0       2
 220454
Median: 10689    13425   10      0       2       440907
75%-tile:        13399   13425   20      0       3
 661360
97.5%-tile:      13425   13425   36      0       5
 859767
Maximum:13425    13425   257     9       8       881812
Mean:    8396.6  9167.03 12.6734 0.0459191       2.41531
# of unique seqs:       761561
total # of seqs:        881812
```

Table 4. Taxonomy output of Classify.seq

taxlevel	rankID	taxon	Daughter levels
taxonomy	total	Lacto bacillus1	Lacto bacillus2
Root	10888	2744	8144
Bacteria;Bacteria_unclassified; Bacteria_unclassified; Bacteria_unclassified; Bacteria_unclassified; Bacteria_unclassified;	9256	2288	6968
unknown;unknown_unclassified; unknown_unclassified; unknown_unclassified; unknown_unclassified; unknown_unclassified;	1632	456	1176

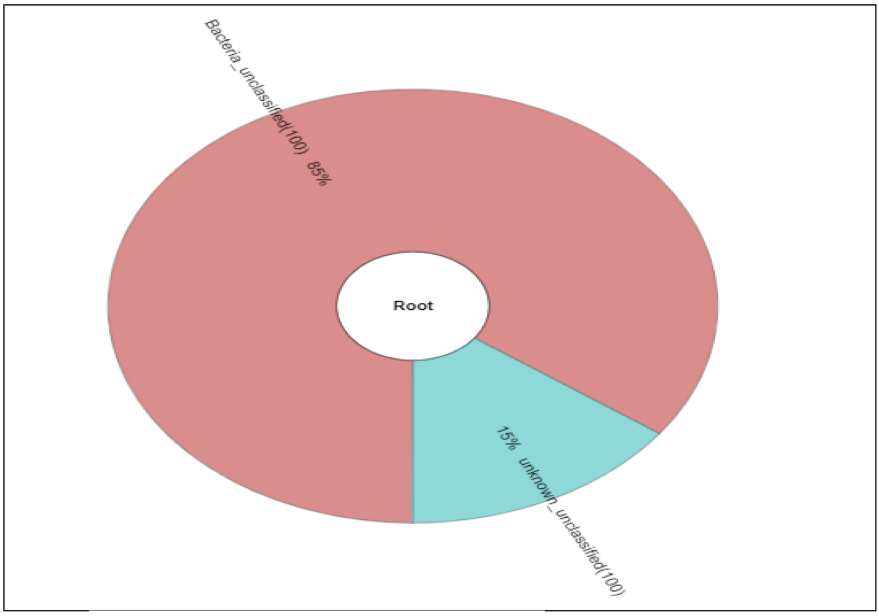

Fig. 1. Krona pie chart visualization of the taxonomy

Structure Based Drug Designing of Bacterial Vaginosis (BV) Disease

Since, Bacterial vaginosis (BV) is a vaginal infection caused by bacteria, we further go ahead towards designing novel drug for the disease. The gene receptors corresponding to BV are taken from NCBI for our work (Table 7).

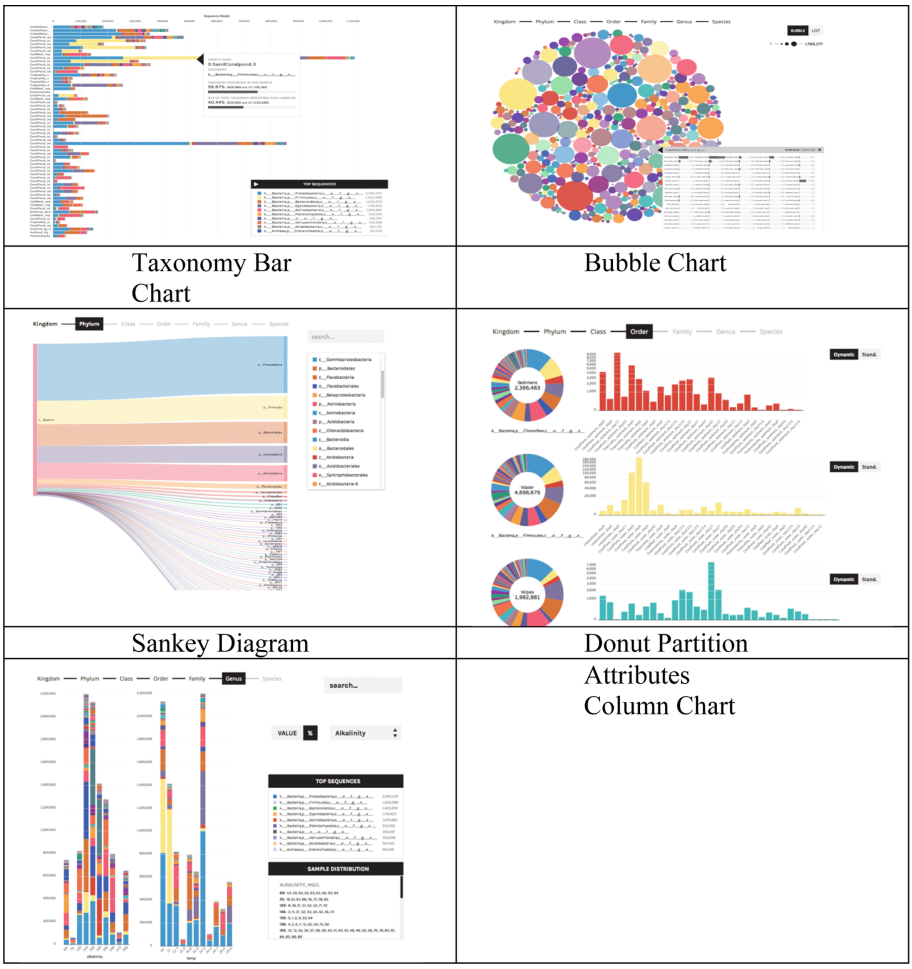

Fig. 2. Pinch visualization of taxonomy data

Abbreviations of Genes:

1. LAP3 - Leucine Aminopeptidase 3
2. PEPD - Peptidase D

Homology Modeling

Homology modeling of the above receptors are done using SWISS-MODEL server. The receptor model and corresponding ramachandran plot results are given in Fig. 3. Template used for modeling is given in Table 7.

Ayurvedic Medicinal plants Gokshura [*Tribulus terrestris*], Nirgundi [*Vitex negundo*], Punarnava [*Boerhaavia diffusa*] and Jatamamsi [*Nardostachys jatamansi*]

Table 5(a). Pathways and their abundance file of SRR011131

# Pathway	humann2
# Pathway	humann2
UNMAPPED	73289.8295612541
UNINTEGRATED	64728.3938932626
UDPNAGSYN-PWY: UDP-N-acetyl-D-glucosamine biosynthesis I	112.5805846312
PWY-7220: adenosine deoxyribonucleotides de novo biosynthesis II	111.1329769439
PWY-7222: guanosine deoxyribonucleotides de novo biosynthesis II	111.1329769439
PWY-5100: pyruvate fermentation to acetate and lactate II	104.6545774222
PWY-7229: superpathway of adenosine nucleotides de novo biosynthesis I	104.4265905210
PWY-7219: adenosine ribonucleotides de novo biosynthesis	100.5920241247
PWY-6126: superpathway of adenosine nucleotides de novo biosynthesis II	98.5710860054
PWY-6386: UDP-N-acetylmuramoyl-pentapeptide biosynthesis II (lysine-containing)	89.8729508007
PWY-621: sucrose degradation III (sucrose invertase)	83.8302023527
PWY-3841: folate transformations II	83.7964388457
GLCMANNANAUT-PWY: superpathway of N-acetylglucosamine, N-acetylmannosamine and N-acetylneuraminate degradation	82.3669178617
COA-PWY-1: coenzyme A biosynthesis II (mammalian)	81.0126259344
PWY-6387: UDP-N-acetylmuramoyl-pentapeptide biosynthesis I (meso-diaminopimelate containing)	80.9328019441
PWY-7221: guanosine ribonucleotides de novo biosynthesis	79.5703173520
PEPTIDOGLYCANSYN-PWY: peptidoglycan biosynthesis I (meso-diaminopimelate containing)	78.2285227543
PWY-6527: stachyose degradation	77.0193132687
PWY-5384: sucrose degradation IV (sucrose phosphorylase)	74.3918043240
PWY-2941: L-lysine biosynthesis II	70.8786963902
PWY-1042: glycolysis IV (plant cytosol)	70.5072210475
PWY-2942: L-lysine biosynthesis III	70.0327634441
GLYCOLYSIS: glycolysis I (from glucose 6-phosphate)	69.2378866007
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose)	66.5376393077
PWY-5686: UMP biosynthesis	65.9584602111
PWY-6317: galactose degradation I (Leloir pathway)	61.9357626169
PWY66–422: D-galactose degradation V (Leloir pathway)	61.9357626169

(*continued*)

Table 5(a). (*continued*)

# Pathway	humann2
PWY66–400: glycolysis VI (metazoan)	61.6522781787
PWY-6151: S-adenosyl-L-methionine cycle I	57.5787595105
PWY-5097: L-lysine biosynthesis VI	55.2908233975
PWY-5667: CDP-diacylglycerol biosynthesis I	27.6416707546
PWY0–1319: CDP-diacylglycerol biosynthesis II	27.6416707546
PWY-5484: glycolysis II (from fructose 6-phosphate)	25.3893258350
TRNA-CHARGING-PWY: tRNA charging	13.2046345369
PWY4FS-7: phosphatidylglycerol biosynthesis I (plastidic)	11.6516504445
PWY4FS-8: phosphatidylglycerol biosynthesis II (non-plastidic)	11.6516504445
ANAEROFRUCAT-PWY: homolactic fermentation	11.5188571644
CALVIN-PWY: Calvin-Benson-Bassham cycle	11.0543406080
PWY-6385: peptidoglycan biosynthesis III (mycobacteria)	9.7975172616
HEXITOLDEGSUPER-PWY: superpathway of hexitol degradation (bacteria)	8.9704383751
PWY-6471: peptidoglycan biosynthesis IV (Enterococcus faecium)	8.7790178439
P441-PWY: superpathway of N-acetylneuraminate degradation	7.1036588677
THRESYN-PWY: superpathway of L-threonine biosynthesis	6.6927872442
PHOSLIPSYN-PWY: superpathway of phospholipid biosynthesis I (bacteria)	6.4743203053
FERMENTATION-PWY: mixed acid fermentation	5.3630548487
GLUCONEO-PWY: gluconeogenesis I	5.0180392541
NONOXIPENT-PWY: pentose phosphate pathway (non-oxidative branch)	4.9895315903
ASPASN-PWY: superpathway of L-aspartate and L-asparagine biosynthesis	4.8875823323
PWY0–1479: tRNA processing	4.8168118491
PWY-7199: pyrimidine deoxyribonucleosides salvage	4.4284215390
PWY-4981: L-proline biosynthesis II (from arginine)	3.8006606994
PWY-724: superpathway of L-lysine, L-threonine and L-methionine biosynthesis II	3.6984681388
PWY0–1586: peptidoglycan maturation (meso-diaminopimelate containing)	3.3167669052
PWY-6124: inosine-5'-phosphate biosynthesis II	3.2002469093
PWY-6122: 5-aminoimidazole ribonucleotide biosynthesis II	3.1726641109

(*continued*)

Table 5(a). (*continued*)

# Pathway	humann2
PWY-6277: superpathway of 5-aminoimidazole ribonucleotide biosynthesis	3.1726641109
PWY-6901: superpathway of glucose and xylose degradation	3.1207222732
PWY-6609: adenine and adenosine salvage III	3.0522701259
PENTOSE-P-PWY: pentose phosphate pathway	3.0214885113
PWY-5913: TCA cycle VI (obligate autotrophs)	2.9695902345
PWY-5695: urate biosynthesis/inosine 5'-phosphate degradation	2.9192588126
PWY-7111: pyruvate fermentation to isobutanol (engineered)	2.7727852823
PWY-6121: 5-aminoimidazole ribonucleotide biosynthesis I	2.7590629291
PWY-6892: thiazole biosynthesis I (E. coli)	2.6709323823
PWY-7117: C4 photosynthetic carbon assimilation cycle, PEPCK type	2.4829652250
PWY-6608: guanosine nucleotides degradation III	2.4565120281
PWY-6891: thiazole biosynthesis II (Bacillus)	2.2995526220
PWY-5723: Rubisco shunt	2.2641148147
OANTIGEN-PWY: O-antigen building blocks biosynthesis (E. coli)	2.2553442744
PWY-3001: superpathway of L-isoleucine biosynthesis I	2.2075630917
P185-PWY: formaldehyde assimilation III (dihydroxyacetone cycle)	2.1888694771
PWY0–1296: purine ribonucleosides degradation	2.1714593387
PWY-4041: γ-glutamyl cycle	2.1164567689
PWY-5138: unsaturated, even numbered fatty acid β-oxidation	2.0268953329
PWY-7357: thiamin formation from pyrithiamine and oxythiamine (yeast)	2.0252027667
PWY-6737: starch degradation V	1.9083529303
PWY-5989: stearate biosynthesis II (bacteria and plants)	1.8971119608
1CMET2-PWY: N10-formyl-tetrahydrofolate biosynthesis	1.8780362349
PWY-6803: phosphatidylcholine acyl editing	1.8684021386
PWY-6123: inosine-5'-phosphate biosynthesis I	1.8610281156
FAO-PWY: fatty acid β-oxidation I	1.7997530499
VALSYN-PWY: L-valine biosynthesis	1.7985043231
THISYNARA-PWY: superpathway of thiamin diphosphate biosynthesis III (eukaryotes)	1.7295333365
THISYN-PWY: superpathway of thiamin diphosphate biosynthesis I	1.7294594209
PWY-5136: fatty acid β-oxidation II (peroxisome)	1.6936319863

(*continued*)

Table 5(a). (*continued*)

# Pathway	humann2
PWY-5973: cis-vaccenate biosynthesis	1.6816508835
PWY-6895: superpathway of thiamin diphosphate biosynthesis II	1.6747372820
COA-PWY: coenzyme A biosynthesis I	1.6645475127
PWY-4242: pantothenate and coenzyme A biosynthesis III	1.6638730103
PWY-5367: petroselinate biosynthesis	1.5454345761
PWY-6897: thiamin salvage II	1.5149815755
PWY0–1061: superpathway of L-alanine biosynthesis	1.4845175809
RIBOSYN2-PWY: flavin biosynthesis I (bacteria and plants)	1.4652452832
GLYOXYLATE-BYPASS: glyoxylate cycle	1.4478093555
PWY-7663: gondoate biosynthesis (anaerobic)	1.3632461609
PWY-1269: CMP-3-deoxy-D-manno-octulosonate biosynthesis I	1.2889190753
PWY-5173: superpathway of acetyl-CoA biosynthesis	1.2874393759
DTDPRHAMSYN-PWY: dTDP-L-rhamnose biosynthesis I	1.2642225032
PWY-7315: dTDP-N-acetylthomosamine biosynthesis	1.2557117889
FASYN-INITIAL-PWY: superpathway of fatty acid biosynthesis initiation (E. coli)	1.2442790103
PWY-5659: GDP-mannose biosynthesis	1.2045435293
COLANSYN-PWY: colanic acid building blocks biosynthesis	1.1819622271
LACTOSECAT-PWY: lactose and galactose degradation I	1.1627906977
NAD-BIOSYNTHESIS-II: NAD salvage pathway II	1.1585795408
PWY0–1533: methylphosphonate degradation I	1.1151379983
PWY-7323: superpathway of GDP-mannose-derived O-antigen building blocks biosynthesis	1.0608653787
PWY-3781: aerobic respiration I (cytochrome c)	0.9909233478
ORNDEG-PWY: superpathway of ornithine degradation	0.9867897688
SO4ASSIM-PWY: sulfate reduction I (assimilatory)	0.9826879308
FUCCAT-PWY: fucose degradation	0.9288285150
PWY-4702: phytate degradation I	0.7818608288
P42-PWY: incomplete reductive TCA cycle	0.6522137798
PWY-5104: L-isoleucine biosynthesis IV	0.5193688068
PWY-6837: fatty acid beta-oxidation V (unsaturated, odd number, di-isomerase-dependent)	0.4699248120

Ghambari [*Gmelina arborea*] are traditionally used to treat many bacterial diseases [24]. The potency of their phytocompounds in treating BV is studied here.

Table 5(b). Pathways and their abundance file of SRR011132

# Pathway	humann2
# Pathway	humann2
UNMAPPED	44811.6942727169
UNINTEGRATED	253211.1152254631
PWY-7220: adenosine deoxyribonucleotides de novo biosynthesis II	284.0306128199
PWY-7222: guanosine deoxyribonucleotides de novo biosynthesis II	284.0306128199
PWY-7229: superpathway of adenosine nucleotides de novo biosynthesis I	277.3477394724
PWY-6126: superpathway of adenosine nucleotides de novo biosynthesis II	271.1262679510
PWY-7219: adenosine ribonucleotides de novo biosynthesis	263.3833651134
GLCMANNANAUT-PWY: superpathway of N-acetylglucosamine, N-acetylmannosamine and N-acetylneuraminate degradation	225.9196615960
PWY-5100: pyruvate fermentation to acetate and lactate II	209.8002175076
PWY-621: sucrose degradation III (sucrose invertase)	202.2746816167
PWY-5384: sucrose degradation IV (sucrose phosphorylase)	190.9573972943
PWY-7221: guanosine ribonucleotides de novo biosynthesis	174.5707505994
UDPNAGSYN-PWY: UDP-N-acetyl-D-glucosamine biosynthesis I	167.6777143736
PWY-6386: UDP-N-acetylmuramoyl-pentapeptide biosynthesis II (lysine-containing)	164.4856349433
COA-PWY-1: coenzyme A biosynthesis II (mammalian)	158.3582975408
PWY-2941: L-lysine biosynthesis II	148.4616670111
PWY-3841: folate transformations II	148.4060659091
PWY-2942: L-lysine biosynthesis III	140.2578853327
PWY-1042: glycolysis IV (plant cytosol)	139.4206921983
PWY-6317: galactose degradation I (Leloir pathway)	138.7709289021
PWY66–422: D-galactose degradation V (Leloir pathway)	138.7709289021
PWY-5097: L-lysine biosynthesis VI	138.3330930682
PWY-5686: UMP biosynthesis	138.1044386354
PWY-6527: stachyose degradation	137.7834036327
PWY-6151: S-adenosyl-L-methionine cycle I	136.4738490516
PEPTIDOGLYCANSYN-PWY: peptidoglycan biosynthesis I (meso-diaminopimelate containing)	135.3321118439
GLYCOLYSIS: glycolysis I (from glucose 6-phosphate)	133.6039727078

(*continued*)

As per Lipinski's rule of five [ADME (Adsorption, distribution and metabolism extraction)] we check the drug likeliness of the above phytocompounds (Tables 8, 9, 10, 11, 12 and 13).

Table 5(b). (*continued*)

# Pathway	humann2
PWY-6387: UDP-N-acetylmuramoyl-pentapeptide biosynthesis I (meso-diaminopimelate containing)	132.9584724951
ANAGLYCOLYSIS-PWY: glycolysis III (from glucose)	128.1722354777
PWY0–1319: CDP-diacylglycerol biosynthesis II	123.8029124029
PWY-5667: CDP-diacylglycerol biosynthesis I	123.8029124029
PWY66–400: glycolysis VI (metazoan)	117.488292209

Table 6(a). Normalized the gene family abundances table of SRR011131 sequence

# Gene Family	humann2-RELAB
# Gene Family	humann2-RELAB
UNMAPPED	0.521182
UniRef50_D5GZK1: Transposase	0.00604246
UniRef50_P35880: Transposase for insertion sequence element IS1201	0.00543639
UniRef50_V6PHI4	0.0049563
UniRef50_unknown	0.00477963
UniRef50_F0NT85: Transposase, IS4 family	0.00321987
UniRef50_W5XI55: Transposase ISLhe15	0.00227109
UniRef50_J3X9M7: Transposase	0.00214584
UniRef50_U2IUJ3	0.00212627
UniRef50_W5XM90: Transposase-like protein	0.0018506
UniRef50_V6NZR9: DNA-binding transcriptional activator CusR	0.00184599
UniRef50_T5LRS1	0.00160632
UniRef50_V6PC60: Cytochrome D ubiquinol oxidase subunit I (Fragment)	0.00148898
UniRef50_S5NEE7: Transposase	0.00142764
UniRef50_H1DWM1: Transaldolase	0.00133821
UniRef50_F7SG55	0.00133249
UniRef50_E4SXC3: Transposase DDE domain	0.00121493
UniRef50_F6CCR0	0.00117058
UniRef50_F8DRE0: Transposase, IS116/IS110/IS902 family	0.00112797
UniRef50_F2LZ32: IS4 family transposase	0.00112379

(*continued*)

Table 6(a). (*continued*)

# Gene Family	humann2-RELAB
UniRef50_E4SIV9: Minor head protein	0.00103507
UniRef50_C2KCD6	0.0010309
UniRef50_D4FFP8	0.00102892
UniRef50_F0NRX0: Transposase	0.00101604
UniRef50_F0TFF9: Lj965 prophage replication protein	0.000961651
UniRef50_W4JLT0	0.000953178
UniRef50_S1SIV9: Transposase	0.00094765
UniRef50_F0TFG1	0.00094236
UniRef50_F0TE53: Transposase	0.000919474
UniRef50_E4SIW6	0.000915181
UniRef50_F0TFG0: Prophage replication protein	0.000914663
UniRef50_E4SIT0	0.000910392

Table 6(b). Normalized the gene family abundances table of SRR011132 sequence

# Gene Family	humann2-RELAB
# Gene Family	humann2-RELAB
UNMAPPED	0.147797
UniRef50_R7QHL2: Stackhouse genomic scaffold, scaffold_290	0.0133392
UniRef50_C7Y3U8	0.0117265
UniRef50_C2KCD4	0.0116592
UniRef50_K1MRL1	0.0100802
UniRef50_F6CF20	0.00991496
UniRef50_W4JKN7	0.00964531
UniRef50_K4AUT0	0.00913137
UniRef50_D5GZN1: Conserved protein	0.00879377
UniRef50_D5GZK1: Transposase	0.00826527
UniRef50_D0DH41	0.00806061
UniRef50_P35880: Transposase for insertion sequence element IS1201	0.00718709
UniRef50_A8YWR8	0.00637412
UniRef50_D5GYT1: Acylphosphatase	0.00618791

(*continued*)

Table 6(b). (*continued*)

# Gene Family	humann2-RELAB
UniRef50_F0TFF7	0.00601488
UniRef50_C2KCD7	0.00586622
UniRef50_unknown	0.00550998
UniRef50_C7XJV5: Gram-positive signal peptide protein, YSIRK family	0.00540967
UniRef50_C2EQ18	0.00531702
UniRef50_F0NS54	0.00522739
UniRef50_E4SIT0	0.00522698
UniRef50_F0NT85: Transposase, IS4 family	0.00517426
UniRef50_C7XLI5	0.00510379
UniRef50_E4SIX3	0.00509921
UniRef50_C2KCD2	0.00498789
UniRef50_C7XJX5	0.00483173
UniRef50_E4SIS9	0.00475573
UniRef50_C7Y5G0	0.00464882
UniRef50_F6CBD8	0.00438834
UniRef50_D0DHI0	0.00434667
UniRef50_D0DFM8	0.00428442
UniRef50_F2M3C6	0.00422323
UniRef50_C7XKA4	0.00401693
UniRef50_D5H125: Conserved protein	0.00381214

Table 7. Genes with their NCBI Accession number

Sl. No	Gene Receptors	NCBI Accession Number	Homologous Template
1.	LAP3	CAG33409.1	2J9A
2	PEPD	CAG46470.1	6H2Q

Molecular Docking

Further docking is performed with the receptors in Table 9 with the above phytocompounds. Docking scores, interacting amino acids along with number of interactions are noted in Tables 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25.

As per docking studies it is seen that phytocompounds Kaempferol, Oxalic acid, Angelic acid, Orselol and Chlorozotocin docks with good interactions with the gene receptors involved in BV, i.e., LAP3 and PEPD.

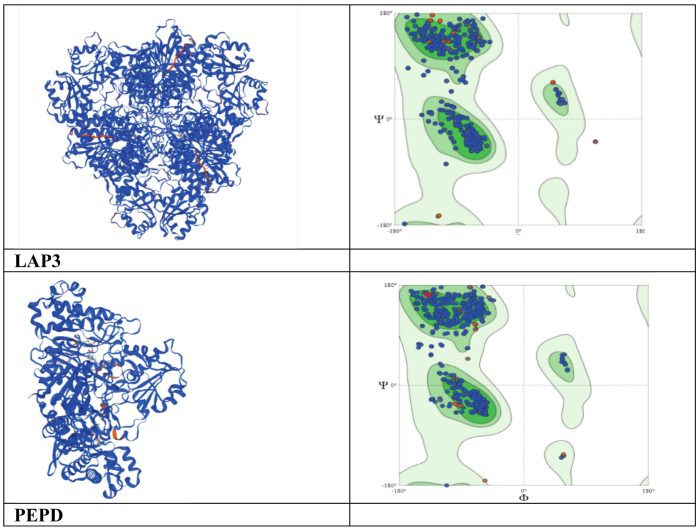

Fig. 3. Swiss-model generated receptor models with their ramachandran plot

Table 8. ADME studies of Gokshura

Sl. No.	Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	Volume
1.	Harmane	2.59	28.68	14	182.23	2	1	0	0	169.42
2.	Harmine	2.63	37.92	16	212.25	3	1	0	1	194.96
3.	Kaempferol	2.17	111.12	21	286.24	6	4	0	1	232.07
4.	Quercetin	1.68	131.35	22	302.24	7	5	0	1	240.08

Table 9. ADME studies of Nirgundi

Sl. No.	Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	Volume
1.	Alanine	-2.69	63.32	6	89.09	3	3	0	1	84.31
2.	Alpha-pinene	3.54	0.00	10	136.24	0	0	0	0	151.81
3.	Camphene	3.33	0.00	10	136.24	0	0	0	0	152.37
4.	Glycine	-2.55	63.32	5	75.07	3	3	0	1	67.73
5.	leucine	-1.38	63.32	9	131.18	3	3	0	3	134.50

(*continued*)

Table 9. (*continued*)

Sl. No.	Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	Volume
6.	luteolin	1.97	111.12	21	286.24	6	4	0	1	232.07
7.	valine	-1.91	63.32	8	117.15	3	3	0	2	117.70
8.	Vanillic acid	1.19	66.76	12	168.15	4	2	0	2	144.61

Table 10. ADME studies of Punarnava

Sl. No.	Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	Volume
1.	Alkaloid	1.67	40.54	19	261.37	3	1	0	0	254.46
2.	Asparagine	-2.81	106.42	9	132.12	5	5	0	3	114.83
3.	Oxalic acid	-1.20	74.60	6	90.03	4	2	0	1	66.64

Table 11. ADME studies of Jatamamsi

Sl. No.	Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	Volume
1.	Angelic acid	1.00	37.30	7	100.12	2	1	0	1	100.17
2.	Angelicin	2.29	43.35	14	186.17	3	0	0	0	154.15
3.	Aristolene	4.84	0.00	15	204.36	0	0	0	0	224.47
4.	Beta-eudesmol	4.01	20.23	16	222.37	1	1	0	1	243.86
5.	Calarene	4.84	0.00	15	204.36	0	0	0	0	224.47
6.	Elemol	4.35	20.23	16	222.37	1	1	0	3	248.59
7.	jatamansin	4.12	65.75	24	328.36	5	0	0	3	298.49
8.	Jatamansinol	2.18	59.67	18	246.26	4	1	0	0	218.00
9.	Jatamansione	1.99	56.52	18	244.25	4	0	0	0	212.14
10.	Orselol	2.68	63.58	18	244.25	4	1	0	1	211.79
11.	Orselone	3.26	43.35	17	226.23	3	0	0	1	198.44
12.	Seselin	3.19	39.45	17	228.25	3	0	0	0	203.77

Table 12. ADME studies of gambhari

Sl. No.	Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	Volume
1.	Apigenin	2.46	90.89	20	270.24	5	3	0	1	224.05
2.	gmelinol	2.26	75.63	29	402.44	7	1	0	6	362.53
3.	Paulowin	2.74	75.63	27	370.36	7	1	0	2	308.20

Table 13. ADME studies of Glycyrrhiza glabra

Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	Volume
Chlorozotocin	-1.48	159.75	20	313.69	10	5	0	9	253.94

Table 14. Docking of LAP3 receptor with Gokshura

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	LAP3	Harmane	-3474	THR-83	1
2.	LAP3	Harmine	-3712	ASP-444	1
3.	LAP3	Kaempferol	-4132	ASP-134, LYS-82	2
4.	LAP3	Quercetin	-3998	ASN-113	1

Table 15. Docking of LAP3 receptor with Nirgundi

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	LAP3	Alanine	-2242	ASN-113	2
2.	LAP3	Glycine	-1852	GLN-111, GLU-110, LEU-434, ARG-431, THR-409	6

(*continued*)

Table 15. (*continued*)

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
3.	LAP3	Leucine	-2674	GLU-110, ASN-113, ARG-431, SER-412	5
4.	LAP3	Luteolin	-3948	GLU-136, LYS-79	2
5.	LAP3	Valine	-2596	GLN-111, LEU-434	3

Table 16. Docking of LAP3 receptor with Punarnava

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	LAP3	Alkaloid	-3686	GLU-136, LYS-82	2
2.	LAP3	Asparagine	-2546	ASN-113, GLN-111, GLU-110, LEU-434	6
3.	LAP3	Oxalic acid	-1856	ASN-113, ARG-431, GLU-110, LEU-434, THR-409	7

Table 17. Docking of LAP3 receptor with Jatamamsi

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	LAP3	Angelic acid	-2542	ARG-431, LEU-434	3
2.	LAP3	Angelicin	-3216	GLU-136, ARG-84	2

(*continued*)

Table 17. (*continued*)

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
3.	LAP3	Beta-eudesmol	-3802	GLU-136	1
4.	LAP3	Elemol	-4040	THR-83	1
5.	LAP3	Jatamansin	-4710	LYS-82	1
6.	LAP3	Jatamansinol	-3816	LYS-82	1
7.	LAP3	Jatamansione	-3810	LYS-82	1
8.	LAP3	Orselol	-3938	ARG-457, GLY-394	2
9.	LAP3	Orselone	-3752	THR-83	1
10.	LAP3	Seselin	-3946	THR-393	1

Table 18. Docking of LAP3 receptor with Ghambari

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	LAP3	Apigenin	-3970	THR-83, TYR-87	2
2.	LAP3	Gmelinol	-5660	TYR-87, LYS-82	2
3.	LAP3	Paulowin	-5226	ARG-84, THR-85, TYR-87	3

Table 19. Docking of LAP3 receptor with Glycyrrhiza glabra

RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
LAP3	Chlorozotocin	-4100	ALA-395, ASN-362, ASP-364, ARG-368, LYS-282, LEU-392	9

Table 20. Docking of PEPD receptor with Gokshura

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	PEPD	Harmane	-3772	GLU-412	1
2.	PEPD	Harmine	-3940	HIS-370, HIS-377	2
3.	PEPD	Kaempferol	-4424	ASP-60, GLU-452, GLY-62, HIS-370	4
4.	PEPD	Quercetin	-4286	ASP-60, ASP-287, GLY-62, HIS-370	4

Table 21. Docking of PEPD receptor with Nirgundi

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	PEPD	Alanine	-2084	ARG-398, GLU-412, GLY-367, HIS-366	5
2.	PEPD	Glycine	-1760	ASP-334, PRO-392	3
3.	PEPD	Leucine	-2812	PHE-65	1
4.	PEPD	Luteolin	-4050	GLU-412, GLY-62, HIS-255	3
5.	PEPD	Valine	-2566	HIS-104	1

Table 22. Docking of PEPD receptor with Punarnava

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	PEPD	Alkaloid	-4542	CYS-283, TYR-282	2
2.	PEPD	Asparagine	-2614	PRO-80	1
3.	PEPD	Oxalic acid	-1794	ARG-331, ASP-334, GLU-391, HIS-358, PRO-392, VAL-355	7

Table 23. Docking of PEPD receptor with Jatamamsi

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	PEPD	Angelic acid	-2646	ARG-398, HIS-255, HIS-377	3
2.	PEPD	Angelicin	-3634	GLU-280, TYR-282	2
3.	PEPD	Beta-eudesmol	-4218	HIS-255	1
4.	PEPD	Elemol	-4092	ARG-237	1
5.	PEPD	Jatamansin	-4982	ARG-398, ASP-60, HIS-255	3
6.	PEPD	Jatamansinol	-4200	ASP-276, THR-242	3
7.	PEPD	Jatamansione	-4164	ASP-276, THR-242	2
8.	PEPD	Orselol	-4158	GLU-53, HIS-104, MET-108	3
9.	PEPD	Orselone	-4002	ARG-398	2
10.	PEPD	Seselin	-4174	ARG-398	2

Table 24. Docking of PEPD receptor with Ghambari

Sl. No.	RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
1.	PEPD	Apigenin	-4090	GLY-62	1
2.	PEPD	Gmelinol	-5682	THR-54, ARG-56	2
3.	PEPD	Paulowin	-5236	ARG-398	1

Table 25. Docking of PEPD receptor with Glycyrrhiza glabra

RECEPTOR	COMPOUND	DOCKING SCORE (Kcal/mol)	INTERACTING AMINO ACID	No. OF INTERACTION
PEPD	Chlorozotocin	-4412	ASP-60, GLY-62, THR-242	3

4 Conclusion

The taxonomy and functional information of BV microbiome are identified. Again, as per docking studies and ADME analysis it is seen that phytocompounds Kaempferol, Oxalic acid, Angelic acid, Orselol and Chlorozotocin can be potential ligands for the receptors implicated in BV. Further, *in-vitro* and *in-vivo* studies can be done on the above phytocompounds to establish their potential as drugs in treating BV.

References

1. Sobel, J. D. (2000). Bacterial vaginosis. *Annual review of medicine*, *51*(1), 349-356.
2. Spiegel, C. A. (1991). Bacterial vaginosis. *Clinical microbiology reviews*, *4*(4), 485-502.
3. Hill, G. B. (1993). The microbiology of bacterial vaginosis. *American journal of obstetrics and gynecology*, *169*(2), 450-454.
4. Livengood III, C. H. (2009). Bacterial vaginosis: an overview for 2009. *Reviews in obstetrics and Gynecology*, *2*(1), 28.
5. Eschenbach, D. A. (1993). History and review of bacterial vaginosis. *American journal of obstetrics and gynecology*, *169*(2), 441-445.
6. Hallén, A., Jarstrand, C. O. N. N. I. E., & Påhlson, C. (1992). Treatment of bacterial vaginosis with lactobacilli. *Sexually transmitted diseases*, *19* (3), 146-148.
7. Swidsinski, A., Mendling, W., Loening-Baucke, V., Ladhoff, A., Swidsinski, S., Hale, L. P., & Lochs, H. (2005). Adherent biofilms in bacterial vaginosis. *Obstetrics & Gynecology*, *106*(5 Part 1), 1013–1023.
8. Turovskiy, Y., Sutyak Noll, K., &Chikindas, M. L. (2011). The aetiology of bacterial vaginosis. *Journal of applied microbiology*, *110*(5), 1105-1128.
9. Verstraelen, H., & Verhelst, R. (2009). Bacterial vaginosis: an update on diagnosis and treatment. *Expert review of anti-infective therapy*, *7*(9), 1109-1124.
10. BISWAS, M. K. (1993). Bacterial vaginosis. *Clinical obstetrics and gynecology*, *36*(1), 166-176.
11. Hay, P. (2014). Bacterial vaginosis. *Medicine*, *42*(7), 359-363.
12. Blankenberg, D., & Hillman-Jackson, J. (2014). Analysis of next-generation sequencing data using Galaxy. In *Stem cell transcriptional networks* (pp. 21–43). Humana Press, New York, NY.
13. Okoli, A. C., Agbakoba, N. R., Ezeanya, C. C., Oguejiofor, C. B., &Anukam, K. C. (2019). Comparative abundance and functional biomarkers of the vaginal and gut microbiome of Nigerian women with bacterial vaginosis: A study with 16S rRNA metagenomics. *J Med Lab Sci*, *29*, 1-26.
14. Hao, S. L., Castaneda, G. R., Cohen, C. R., Hemmerling, A., & Crawford, E. D. (2020). Metagenomic next generation sequencing analysis of vaginal microbiome composition in patients with bacterial vaginosis treated with lactin-v (Lactobacillus crispatus CTV-05) versus placebo. *American Journal of Obstetrics & Gynecology*, *223*(6), 967-968.
15. Srinivasan, S., & Fredricks, D. N. (2008). The human vaginal bacterial biota and bacterial vaginosis. *Interdisciplinary perspectives on infectious diseases*, *2008*.
16. Saskia Hiltemann, Bérénice Batut, 2020 Analyses of metagenomics data - The global picture (Galaxy Training Materials). https://training.galaxyproject.org/training-material/topics/metagenomics/tutorials/general-tutorial/tutorial.html Online; accessed Tue Aug 17 2021
17. Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/aem.01541-09

18. Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1). https://doi.org/10.1186/1471-2105-12-385
19. Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., … Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, *12*(10), 902–903.
20. Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., … Huttenhower, C. (2012). Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLoS Computational Biology*, *8*(6), e1002358. https://doi.org/10.1371/journal.pcbi.1002358
21. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, A P de Beer T, Rempfer C, Bordoli L, Lepore R and Schwede T, (2018), SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res.; 46(Web Server issue): W296–W303.
22. https://www.molinspiration.com, Slovensky Grob, Slovakia
23. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl. Acids. Res. 33: W363-367, 2005.
24. Hegde PL & Harini A, 2014, A text book of Dravyaguna Vijnana, Chaukhambha publications

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.