Establishing the Taxa with Phylogenetic Profile and *in-silico Ayurvedic* Remedy of Human Oropharynx Microbiome

B. Yeswanthi Kumari[1*], Shylesh Murthy IA[2] and Preenon Bagchi[3]

[1]Padmashree Institute of Management and Sciences, Bengaluru, India. Corresponding author: Email: yeswanthikumari1999@gmail.com

[2]Junior Researcher, Vasishth Academy of Advanced Studies and Research, Bengaluru, India.

[3]Institute of Biosciences and Technology, MGM University, Bengaluru, India.

Abstract:

Fusobacterium, a Gram-negative bacterium, is the cause of oropharyngeal cancer, a type of head and neck cancer. We have performed metagenomics analysis on the Human Oropharynx Microbiome and identify the taxonomic information and gene family information. Further, we have identified the receptors involved in Oropharynx and performed computer-aided drug design using the phytocompounds from the ayurvedic medicinal herbs.

Keywords: oropharyngeal, microbiome, fusobacterium, galaxy, metagenomics, next-generation sequencing, Ayurvedic medicinal herbs, Operational Taxonomic Units, taxonomy, modelling, Lipinski's rule of five, docking

INTRODUCTION

Fusobacterium is one among the causes of oropharyngeal cancer, a type of head and neck cancer in which cancer cells are found within oropharynx [1, 2, 3, 4]. Fusobacterium is a Gram-negative bacterium which belongs to class Fusobacteria [5]. This bacterium doesn't produce spores [6, 7, 8]. It is non-motile and anaerobic [9]. The individual cells of this bacteriam are slender, rod shaped bacilli with pointed ends and has a G-C content of 27-28% [7]. This bacterium being an oral bacterium is found only be found in the mouth cavity of mammals [10, 11]. It is known for causing typical dental plaque on human teeth as well as involvement in periodontal diseases and invasive human infections of the head and neck, chest, lung, liver and abdomen [1, 12].

Metagenomics is the study of the metagenome, i.e., the study of microbes in their natural living environment [13-15]. This study examines the genetic and genomic composition of an entire organism (microbiome), including each of the microbes that exist and co-exist with it [13-15]. This study is carried out to provide information on the microbial ecology

R. Somashekhar et al. (eds.), *Proceedings of the International Conference on Advances in Nano-Neuro-Bio-Quantum (ICAN 2023)*, Advances in Intelligent Systems Research 69,
https://doi.org/10.2991/978-94-6463-294-1_19

and biodiversity [13-15]. It is hoped that increased understanding of the nature of microbes in the environment, i.e., microbiome could have a significant impact on other sciences and research areas, such as medicine, biology, biotechnology and ecology [13-15].

MATERIALS AND METHODS

Galaxy tutorial by Hiltemann S and Batut B, 2020 Analyses of metagenomics data - The global picture is used to analyze human oropharynx microbiome [16]. Fusobacterium' fasta sequences ERR4058556 and ERR4058557 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 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 human oropharynx microbiome, 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].

RESULTS AND DISCUSSION

Fusobacterium' fasta sequences ERR4058556 and ERR4058557 data summary as per summary.seq is given in Table 1.

Table 1: summary.seq output of fasta sequences ERR4058556 and ERR4058557

Start	End	NBases	Ambigs	Polymer	NumSeqs

Minimum:	1	250	250	0	3	1
2.5%-tile:	1	250	250	0	3	2486
25%-tile:	1	250	250	0	4	24857
Median:	1	250	250	0	4	49713
75%-tile:	1	250	250	0	5	74569
97.5%-tile:	1	250	250	32	8	96939
Maximum:	1	250	250	94	227	99424
Mean: 1	250	250	1.30224	4.63251		
# of Seqs:	99424					

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.

Table 2: Align.seq output

QueryName	QueryLength	TemplateName	TemplateLength	SearchMethod	SearchScore	AlignmentMethod	QueryStart	QueryEnd	TemplateStart	TemplateEnd	PairwiseAlignmentLength	GapsInQuery	GapsInTemplate	LongestInsert	SimBtwnQuery&Template
ERR405856.1.1	250	AF491832.1	293	kmer	90.12	needleman	1	247	47	293	247	0	0	0	98.79
ERR405856.1.2	250	AF491832.1	293	kmer	71.19	needleman	1	250	1	250	250	0	0	0	96.00
ERR405856.2.1	250	AF132273.1	293	kmer	99.18	needleman	3	250	1	248	248	0	0	0	100.00
ERR405856.2.2	250	AF132273.1	293	kmer	99.59	needleman	1	250	44	293	250	0	0	0	99.60
ERR405856.3.1	250	AJ408992.1	293	kmer	95.47	needleman	1	248	46	293	248	0	0	0	99.19
ERR405856.3.2	250	AJ408992.1	293	kmer	95.47	needleman	3	250	1	248	248	0	0	0	99.19
ERR405856.4.1	250	AB114225.1	294	kmer	79.42	needleman	4	250	1	247	247	0	0	0	96.36
ERR405856.4.2	250	AY212760.1	293	kmer	72.43	needleman	1	245	49	293	245	0	0	0	93.47
ERR405856.5.1	250	AB267809.1	292	kmer	68.31	needleman	1	249	45	292	249	0	1	0	93.57
ERR405856.5.2	250	AB267809.1	292	kmer	53.09	needleman	1	250	1	249	250	0	1	0	90.00
ERR4	250	AB23	293	kme	93.0	needl	1	24	48	293	246	0	0	0	99.19

QueryName	QueryLength	TemplateName	TemplateLength	SearchMethod	SearchScore	AlignmentMethod	QueryStart	QueryEnd	TemplateStart	TemplateEnd	PairwiseAlignmentLength	GapsInQuery	GapsInTemplate	LongestInsert	SimBtwnQuery&Template
058556.6.1		8928.1		r	0	eman		6							
ERR4058556.6.2	250	AB238928.1	293	kmer	88.89	needleman	1	250	1	250	250	0	0	0	98.80
ERR4058556.7.1	250	AF132273.1	293	kmer	94.65	needleman	1	245	49	293	245	0	0	0	99.59
ERR4058556.7.2	250	AF132273.1	293	kmer	92.18	needleman	3	250	1	248	248	0	0	0	98.39
ERR4058556.8.1	250	AB238928.1	293	kmer	92.18	needleman	1	249	45	293	249	0	0	0	98.80
ERR4058556.8.2	250	AB238928.1	293	kmer	94.24	needleman	3	250	1	248	248	0	0	0	99.19
ERR4058556.9.1	250	AB114225.1	294	kmer	83.95	needleman	1	250	1	250	250	0	0	0	96.40
ERR4058556.9.2	250	Y10030.1	293	kmer	81.48	needleman	1	247	47	293	247	0	0	0	95.95
ERR4058556.10.1	250	AF132273.1	293	kmer	93.00	needleman	2	250	1	249	249	0	0	0	99.20
ERR4058556.10.2	250	AF132273.1	293	kmer	96.30	needleman	1	250	44	293	250	0	0	0	99.20
ERR4058556.11.1	250	AF132251.1	293	kmer	97.94	needleman	1	245	49	293	245	0	0	0	100.00
ERR4058556.11.2	250	AF132251.1	293	kmer	65.43	needleman	1	250	1	250	250	0	0	0	90.80
ERR4058556.12.1	250	AF132251.1	293	kmer	95.47	needleman	1	248	46	293	248	0	0	0	99.19
ERR4058556.12.2	250	AF132251.1	293	kmer	81.48	needleman	4	250	1	247	247	0	0	0	97.17
ERR4058556.13.1	250	AF227870.1	293	kmer	91.36	needleman	2	250	1	249	249	0	0	0	98.80

QueryName	QueryLength	TemplateName	TemplateLength	SearchMethod	SearchScore	AlignmentMethod	QueryStart	QueryEnd	TemplateStart	TemplateEnd	PairwiseAlignmentLength	GapsInQuery	GapsInTemplate	LongestInsert	SimBtwnQuery&Template
ERR4058556.13.2	250	AF227870.1	293	kmer	80.25	needleman	1	250	42	291	250	0	0	0	97.20
ERR4058556.14.1	250	AF132251.1	293	kmer	74.07	needleman	1	245	49	293	245	0	0	0	95.10
ERR4058556.14.2	250	AF132251.1	293	kmer	80.66	needleman	3	250	1	248	248	0	0	0	93.55
ERR4058556.15.1	250	CP000140.1	293	kmer	66.26	needleman	1	250	44	293	250	0	0	0	94.00

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

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	10641	244	0	3	3584
25%-tile:	1	10646	247	0	4	35838
Median:	3067	13424	248	0	4	71676
75%-tile:	3072	13425	249	0	5	107514
97.5%-tile:	3085	13425	250	8	7	139768
Maximum:	13425	13425	250	40	8	143351
Mean: 1580.43	12019.2	246.516	0.758753		4.43573	
# of unique seqs:		97597				
total # of seqs:		143351				

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 Fig. 1 & 2 respectively.

Table 4: taxonomy output of Classify.seq

taxlevel	rankID	Taxon	daughterlevels
taxonomy	total	Combinedprebioticandmicrobialintervention	GutMicrobiome
Root	943	891	52
Bacteria;Actinobacteria;Actinobacteria;Actinobacteria_unclassified;Actinobacteria_unclassified;Actinobacteria_unclassified;	1	1	0
Bacteria;Actinobacteria;Actinobacteria;Bifidobacteriales;Bifidobacteriaceae;Bifidobacterium;	7	6	1
Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Atopobium;	4	4	0
Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Coriobacteriaceae_unclassified;	1	1	0
Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Eggerthella;	27	27	0
Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Gordonibacter;	1	0	1
Bacteria;Bacteria_unclassified;Bacteria_unclassified;Bacteria_unclassified;Bacteria_unclassified;Bacteria_unclassified;	37	35	2

taxlevel	rankID	Taxon	daughter levels
Bacteria;Bacteroidetes;Bacteroidetes_unclassified;Bacteroidetes_unclassified;Bacteroidetes_unclassified;Bacteroidetes_unclassified;	1	1	0
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Anaerorhabdus;	4	4	0
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides;	28	16	12
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidales_unclassified;Bacteroidales_unclassified;	2	2	0
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter;	9	0	9
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Parabacteroides;	5	0	5
Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Porphyromonadaceae_unclassified;	1	0	1
Bacteria;Firmicutes;Bacilli;Bacilli_unclassified;Bacilli_unclassified;Bacilli_unclassified;	9	9	0
Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillaceae;Pediococcus;	1	0	1
Bacteria;Firmicutes;Bacilli;Lactobacillales;Lactobacillales_unclassified;Lactobacillales_unclassified;	20	20	0
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Lactococcus;	128	128	0
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus;	214	214	0
Bacteria;Firmicutes;Clostridia;Clostridiales;Clostridiales_unclassified;Clostridiales_unclassified;	2	2	0
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Dorea;	2	0	2
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Hungatella;	6	0	6
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Lachnospiraceae_unclassified;	54	54	0
Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Robinsoniella;	1	1	0
Bacteria;Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Peptostreptococcaceae_unclassified;	15	15	0
Bacteria;Firmicutes;Clostridia;Clostridiales;Peptostreptococcaceae;Romboutsia;	41	41	0
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Anaerotruncus;	1	0	1
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Faecalibacterium;	2	0	2
Bacteria;Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Ruminococcaceae_unclassified;	4	2	2
Bacteria;Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Clostridium_XVIII;	94	94	0
Bacteria;Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Coprobacillus;	4	0	4
Bacteria;Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Erysipelotrichaceae_unclassified;	13	13	0
Bacteria;Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Faecalicoccus;	75	75	0
Bacteria;Firmicutes;Erysipelotrichia;Erysipelotrichales;Erysipelotrichaceae;Holdemania;	7	4	3
Bacteria;Firmicutes;Firmicutes_unclassified;Firmicutes_unclassified;Firmicutes_unclassified;Firmicutes_unclassified;	26	26	0
Bacteria;Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Veillonella;	84	84	0
Bacteria;Firmicutes;Negativicutes;Selenomonadales;Veillonellaceae;Veillonellaceae_unclassified;	6	6	0
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Enterobacteriaceae_unclassified;	1	1	0
Bacteria;Proteobacteria;Gammaproteobacteria;Gammaproteobacteria_unclassified;Gammaproteobacteria_unclassified;Gammaproteobacteria_unclassified;	1	1	0
unknown;unknown_unclassified;unknown_unclassified;unknown_unclassified;unknown_unclassified;unknown_unclassified;	4	4	0

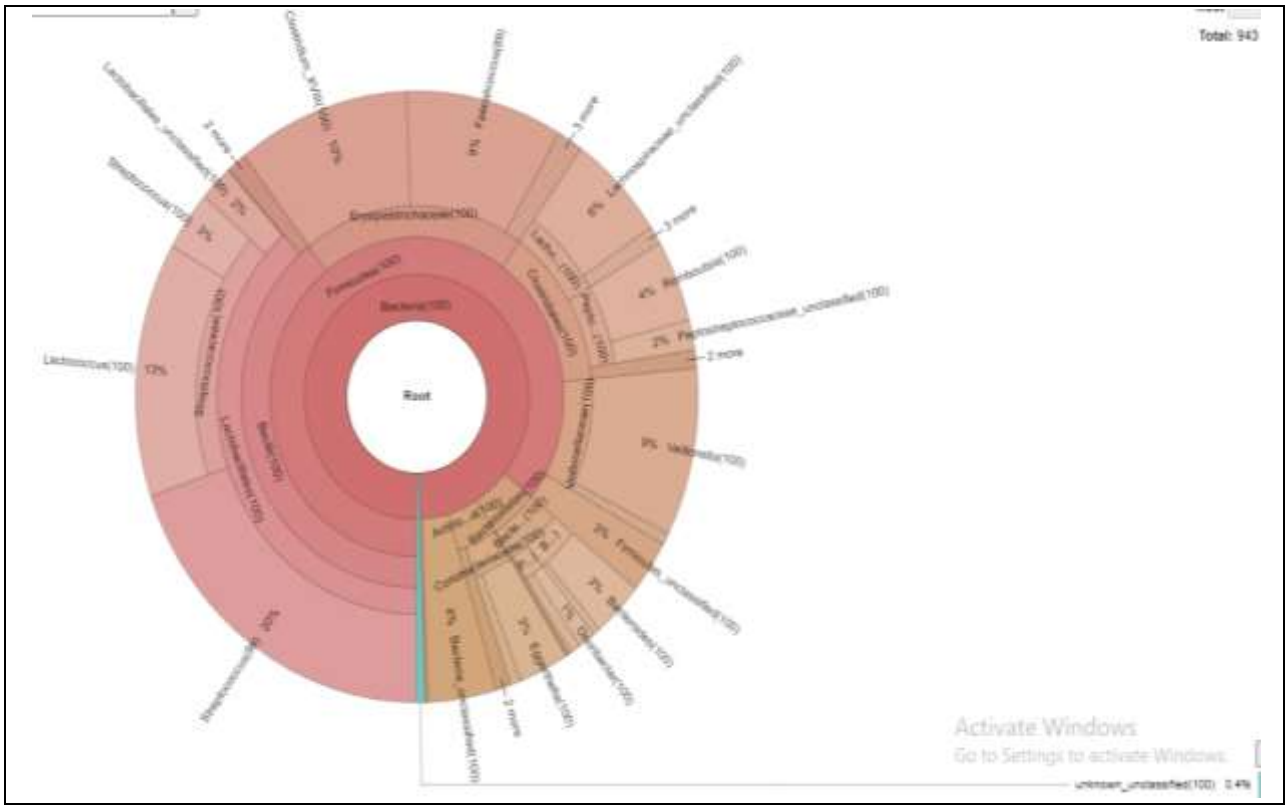
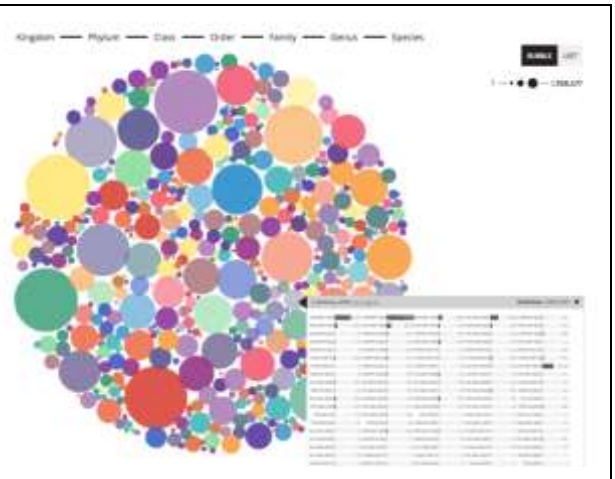


Fig. 1: Krona pie chart visualization of the taxonomy

	
Taxonomy Bar Chart	Bubble Chart

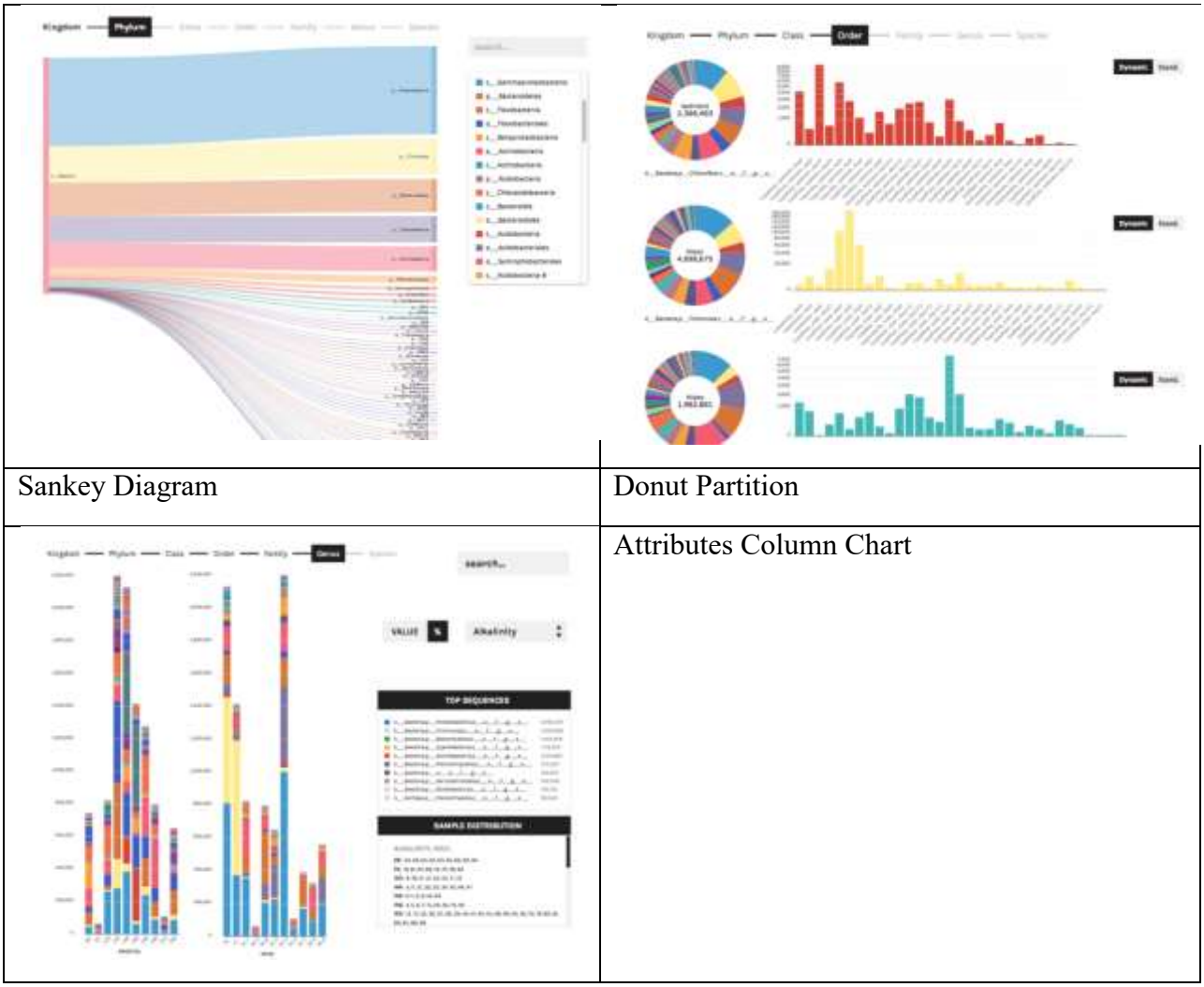


Fig. 2: pinch visualization of taxonomy data

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 ERR4058556 sequence as per the Shotgun metagenomic sequencing and the sequence 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 normalized the gene family abundances table of ERR4058556 sequence is given in table 5. Next the Shotgun metagenomic sequencing of ERR4058557 sequence performed and normalized the gene family abundances table of ERR4058557 is given in table 6 respectively.

Table 5: normalized the gene family abundances table of ERR4058556 sequence

# Gene Family	humann2-RELAB
# Gene Family	humann2-RELAB
UNMAPPED	0.786712

# Gene Family	humann2-RELAB
UniRef50_A0A023GNY3: Putative secreted protein (Fragment)	0.161414
UniRef50_UPI000371FF69: hypothetical protein	0.0350316
UniRef50_UPI0003C39CDF: PREDICTED: putative per-hexamer repeat protein 5-like	0.0120501
UniRef50_A0A023FUY2: Putative secreted protein (Fragment)	0.00163865
UniRef50_K7EWA4	0.000982765
UniRef50_F2RNS9	0.000392399
UniRef50_T1EMK7	0.000349432
UniRef50_U6L5R8	0.000327283
UniRef50_A8QFS0: U88, putative (Fragment)	0.000278946
UniRef50_U6JYQ1	9.66576e-05
UniRef50_A2A5X6: MCG147035	9.54545e-05
UniRef50_M1CWF1	9.39647e-05
UniRef50_Q3KSS4: Epstein-Barr nuclear antigen 1	7.10059e-05
UniRef50_A8P3B5	6.928e-05
UniRef50_C3ZSR4	5.71298e-05
UniRef50_L0A0L5	5.48288e-05
UniRef50_P09789: Glycine-rich cell wall structural protein 1	5.40521e-05
UniRef50_Q1HVF7: Epstein-Barr nuclear antigen 1	4.30274e-05
UniRef50_Q10MG8: Retrotransposon protein, putative, Ty3-gypsy subclass, expressed	4.0686e-05
UniRef50_UPI0002652EE0	2.0965e-05
UniRef50_B3N5G6: GG25361, isoform A	1.87957e-05
UniRef50_UPI00032A2300: PREDICTED: mucin-2-like	1.78989e-05
UniRef50_Q8IMS9: Mucin 96D	1.70628e-05
UniRef50_F2D9R2: Predicted protein	1.5692e-05
UniRef50_A6YP79: Minor ampullate spidroin-like protein (Fragment)	9.86411e-06
UniRef50_B4JUL6: GH15522	8.92867e-06
UniRef50_Q41187: Glycine-rich protein (Fragment)	5.04782e-06
UniRef50_M1BG02	4.90019e-06
UniRef50_B4KDH9: GI22470	4.14957e-06
UniRef50_B4LHU8: GJ11408	4.02261e-06
UniRef50_K7VAT5	2.92509e-06
UniRef50_H9K5U8	2.84704e-06
UniRef50_J3ML95	2.82364e-06
UniRef50_M1CTS0	2.62882e-06
UniRef50_P27483: Glycine-rich cell wall structural protein	1.89548e-06
UniRef50_UPI00032AB6B5: PREDICTED: glycine-rich cell wall structural protein-like isoform X3	1.33023e-06
UniRef50_Q9VR49: Salivary gland secretion 1	1.25634e-06
UniRef50_UPI0004441CAE: PREDICTED: keratin-associated protein 5-5-like	1.21744e-06
UniRef50_Q9VMG7: Mucin 26B	1.08709e-06
UniRef50_Q0C7P7: Predicted protein	8.89485e-07
UniRef50_U5D4P4	7.1303e-07

As per table 5, the gene families identified in ERR4058556 are Putative secreted protein (Fragment), PREDICTED: putative per-hexamer repeat protein 5-like, Putative secreted protein (Fragment), Epstein-Barr nuclear antigen 1, Retrotransposon protein, putative, Ty3-gypsy subclass, expressed, PREDICTED: mucin-2-like, Mucin 96D, Minor ampullate spidroin-like protein (Fragment), Glycine-rich protein (Fragment), Glycine-rich cell wall structural protein, PREDICTED: glycine-rich cell wall structural protein-like isoform X3, Salivary gland secretion 1, PREDICTED: keratin-associated protein 5-5-like and Mucin 26B.

Table 6: normalized the gene family abundances table of ERR4058557 sequence

# Gene Family	humann2-RELAB
# Gene Family	humann2-RELAB
UNMAPPED	0.822831
UniRef50_A0A023GNY3: Putative secreted protein (Fragment)	0.16912
UniRef50_UPI000371FF69: hypothetical protein	0.00213184
UniRef50_A0A023FUY2: Putative secreted protein (Fragment)	0.00127924
UniRef50_K7EWA4	0.000998836
UniRef50_UPI0003C39CDF: PREDICTED: putative per-hexamer repeat protein 5-like	0.000797294
UniRef50_U6L5R8	0.0007266
UniRef50_U6N0F3	0.000483711
UniRef50_F2RNS9	0.000326446
UniRef50_T1EMK7	0.000285784
UniRef50_A8QFS0: U88, putative (Fragment)	0.000234605
UniRef50_A8P3B5	0.000106864
UniRef50_W7JP25: Cyclin-dependent kinases regulatory subunit	0.000102238
UniRef50_C3ZSR4	8.85796e-05
UniRef50_Q1HVF7: Epstein-Barr nuclear antigen 1	8.60986e-05
UniRef50_UPI0003505F44: PREDICTED: keratin-associated protein 9-1-like, partial	6.30656e-05
UniRef50_Q3KSS4: Epstein-Barr nuclear antigen 1	6.14742e-05
UniRef50_F0M9E2	3.22921e-05
UniRef50_M1CWF1	2.73551e-05
UniRef50_B4NLW8: GK10515	2.68805e-05
UniRef50_A0A059N8H5: Triple helix repeat-containing collagen (Fragment)	2.4966e-05
UniRef50_A6YP79: Minor ampullate spidroin-like protein (Fragment)	2.12687e-05
UniRef50_L0A0L5	1.77893e-05
UniRef50_U6GNA4	1.71143e-05
UniRef50_UPI0004441CAE: PREDICTED: keratin-associated protein 5-5-like	1.69223e-05
UniRef50_U6JYQ1	1.44851e-05
UniRef50_UPI0002652EE0	1.19964e-05
UniRef50_B4JUL6: GH15522	1.1873e-05
UniRef50_U6N5L2	1.1483e-05
UniRef50_B3N5G6: GG25361, isoform A	1.05354e-05
UniRef50_U6KGA9	8.16382e-06
UniRef50_Q8IMS9: Mucin 96D	4.49425e-06
UniRef50_Q10MG8: Retrotransposon protein, putative, Ty3-gypsy subclass, expressed	3.49724e-06
UniRef50_UPI00038387E5: PREDICTED: keratin-associated protein 10-4-like, partial	3.26198e-06
UniRef50_G1LP61	1.66327e-06
UniRef50_P27483: Glycine-rich cell wall structural protein	1.57306e-06
UniRef50_T1DKE6: Putative dna repair (Fragment)	1.07765e-06
UniRef50_M1BG02	1.07612e-06
UniRef50_K4MTL7: Minor ampullate spidroin	8.49178e-07
UniRef50_B4LHU8: GJ11408	8.35463e-07
UniRef50_P09789: Glycine-rich cell wall structural protein 1	7.98864e-07
UniRef50_M1CTS0	7.52288e-07
UniRef50_F2D9R2: Predicted protein	7.3476e-07
UniRef50_Q41187: Glycine-rich protein (Fragment)	5.81252e-07
UniRef50_J3ML95	3.95283e-07
UniRef50_UPI00032AB6B5: PREDICTED: glycine-rich cell wall structural protein-like isoform X3	3.71478e-07
UniRef50_B2L207: Minor ampullate spidroin 1-like protein (Fragment)	2.52117e-07
UniRef50_K7VAT5	2.06371e-07

As per table 6, the gene families identified in ERR4058557 are Putative secreted protein (Fragment), PREDICTED: putative per-hexamer repeat protein 5-like, Cyclin-dependent kinases regulatory subunit, Epstein-Barr nuclear antigen 1, PREDICTED: keratin-associated protein 9-1-like, partial, Triple helix repeat-containing collagen (Fragment), Minor ampullate spidroin-like protein (Fragment), PREDICTED: keratin-associated protein 5-5-like, Mucin 96D, Retrotransposon protein, putative, Ty3-gypsy subclass, expressed, PREDICTED: keratin-associated protein 10-4-like, partial, Glycine-rich cell wall structural protein, Putative dna repair (Fragment), Minor ampullate spidroin, PREDICTED: glycine-rich cell wall structural protein-like isoform X3 and Minor ampullate spidroin 1-like protein (Fragment). Further studies can be taken on the gene families identified in table 5 & 6.

Structure based drug designing of Oropharyngeal head and neck cancer disease

Since, human oropharynx is an infection caused by Fusobacterium, we further go ahead towards designing novel drug for the disease. The gene receptors corresponding to oropharynx are taken from NCBI for our work (Table 7).

Table 7: Genes Involved in Oropharynx cancer with their NCBI Accession number

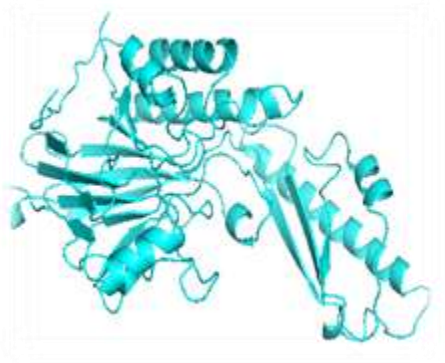
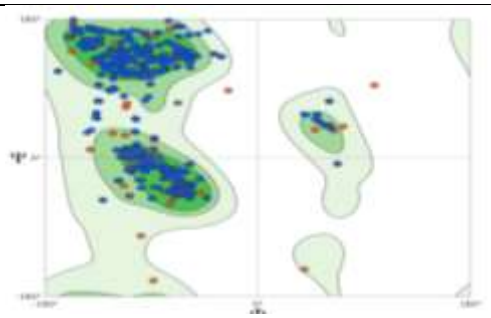
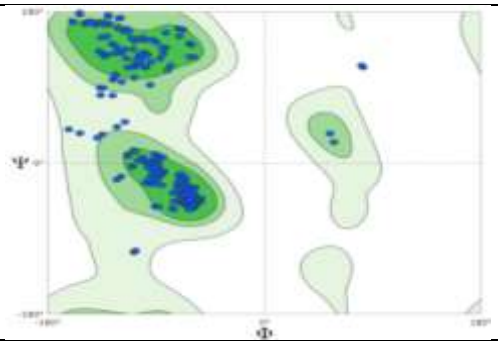
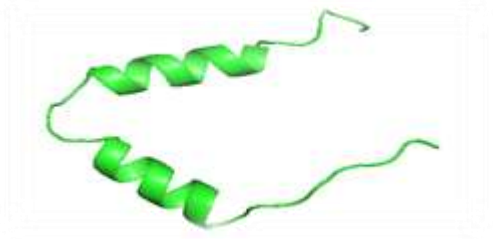
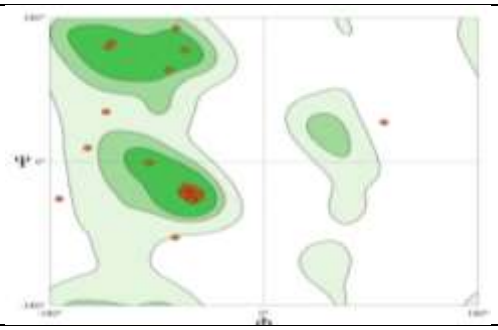
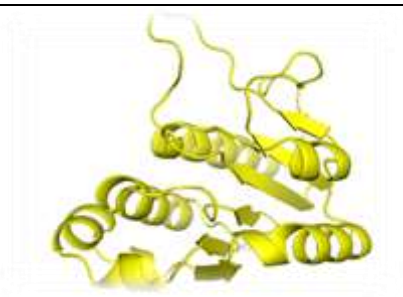
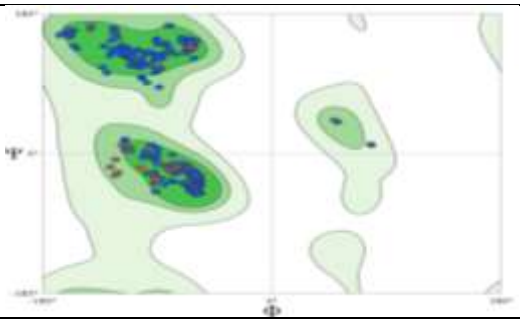
S.no	Genes involved	Accession number	Homologous Template
1	MRE11	NP_001317276.1	3T1IA
2	GSTP1	CAG29357.1	5I6XA
3	CDKN2A	AAH15960.3	1HN3A
4	MIR21	NP_005637.3	2HA8A
5	POLQ	O75417.2	5A9JA

Abbreviations of genes:

1. MRE11 - MRE11 Homolog, Double Strand Break Repair Nuclease
2. GSTP1- Glutathione S-Transferase Pi 1
3. CDKN2A - Cyclin Dependent Kinase Inhibitor 2A
4. MIR21- MicroRNA 21
5. POLQ- DNA Polymerase Theta

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.

	
MRE11	
	
GSTP1	
	
CDKN2A	
	
MIR21	

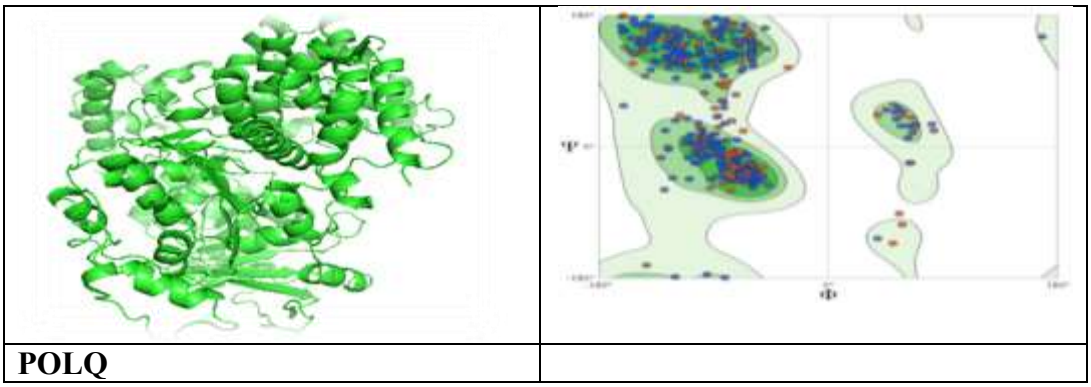


Fig. 3: Swiss-model generated receptor models with their ramachandran plot

Ayurvedic Medicinal plants *Artemisia absinthium*, *Capparis spinosa*, *Cleome drosiferila*, *Glycyrrhiza glabra* and *Harpagophytum procumbens* are traditionally used to treat many diseases including inflammations and cancerous tumors [24]. The potency of their phytocompounds in treating oropharynx is studied here.

As per Lipinski's rule of five [ADME (Adsorption, distribution and metabolism extraction)] we check the drug likeliness of the above phytocompounds using the Molinspiration server. The ADME results are given in Table 8-12.

Table 8: ADME studies of phytocompounds from *Artemisia absinthium*:

S.no	Compound	Mi log P	TPSA	natoms	mw	nON	nOHNH	Nviolations	nROTB	Volume
1	p-cymene	3.90	0.00	10	134.22	0	0	0	1	150.55
2	Isoledene	4.67	0.00	15	204.36	0	0	0	0	224.79
3	Isopulegol acetate	3.35	26.30	14	196.29	2	0	0	3	208.06
4	Camphenol	2.41	20.23	11	152.24	1	1	0	0	160.41
5	Diisoamylene	3.92	17.07	13	182.31	1	0	0	3	208.29
6	Linalool	3.21	20.23	11	154.25	1	1	0	4	175.59
7	n-amyl isovalerate	3.46	26.30	12	172.27	2	0	0	7	191.12
8	2,4-heptadienal	2.05	17.07	8	110.16	1	0	0	3	119.81
9	Myrcene	3.99	0.00	10	136.24	0	0	0	4	162.24
10	Thymol	3.34	20.23	11	150.22	1	1	0	1	158.57

Table 9: ADME studies of phytocompounds from *Capparis spinosa:*

S.no	Compound	Mi log P	TPSA	natoms	mw	nON	nOHNH	Nviolations	nROTB	Volume
1	Apigenin	-2.46	90.89	20	270.24	5	3	0	1	224.05
2	Flazin	-2.79	99.35	23	308.29	6	3	0	3	257.65

3	Capparisin A	-0.12	80.93	18	253.25	6	2	0	2	218.08
4	Kaempferol	2.17	111.12	21	286.24	6	4	0	1	232.07
5	Sakuranetin	2.65	76.00	21	286.28	5	2	0	2	247.79
6	Stachydrine	-5.31	40.13	10	143.19	3	0	0	1	142.62
7	Tetrahydroquinone	2.31	12.03	10	133.19	1	1	0	0	136.01
8	Thevetiaflavone	2.74	79.90	21	284.27	5	2	0	2	241.58

Table 10: ADME studies of phytocompounds from *Cleome drosiferila:*

S.no	Compound	Mi log P	TPSA	natoms	mw	nON	nOHNH	Nviolations	nROTB	Volume
1	1-methylnapthalene	3.54	0.00	11	142.20	0	0	0	0	144.60
2	Acenaphthylene	3.72	0.00	12	152.20	0	0	0	0	144.61
3	Chlorpyrifos	5.16	40.59	18	350.59	4	0	1	6	251.40
4	Anthracene	4.30	0.00	14	178.23	0	0	0	0	172.03
5	Dichlorvos	0.67	44.77	11	220.98	4	0	0	4	155.55
6	Fenchlorphos	4.94	27.70	16	321.55	3	0	0	4	221.95
7	Lindane	3.73	0.00	12	290.83	0	0	0	0	183.97
8	Naphthalene	3.15	0.00	10	128.17	0	0	0	0	128.03

Table 11: ADME studies of phytocompounds from *Harpagophytum procumbens:*

S.no	Compound	Mi log P	TPSA	natoms	mw	Non	nOHNH	Nviolations	Nrotb	Volume
1	1,2-Benzenedicarboxylic acid,bis(2-methylpropyl) ester	3.80	52.61	20	278.35	4	0	0	8	273.48
2	6-Methoxynicotinic acid	0.67	59.42	11	153.14	4	1	0	2	132.43
3	2-(2,5-Dimethoxyphenyl)CYCLOHEX-2-ENONE	2.52	35.54	17	232.28	3	0	0	3	221.10
4	2,6-Dimethyl1-3-aminobenzoquinone	1.66	66.48	11	153.11	3	4	0	0	144.49
5	2-Methoxy-4-vinylphenol	2.13	29.46	11	150.18	2	1	0	2	145.34
6	Cyclohexanecarboxamide,N-furfuryl)	2.36	42.24	15	207.27	3	1	0	3	203.79
7	1S,2S,5R-1,4,4-Trimethyltricyclo[6.3.1.0(2,5)]dodec-8(9)-ene	4.73	0.00	15	204.36	0	0	0	0	224.68

Table 12: ADME studies of phytocompounds from *Glycyrrhiza glabra :*

S.no	Compound	Mi log P	TPSA	natoms	mw	nON	nOHNH	Nviolations	Nrotb	Volume

1.	1-methoxyficifolinol	6.71	68.16	31	422.52	5	2	1	5	398.46
2	Alpha terpineol	2.60	20.23	11	154.25	1	1	0	1	170.65
3	Chlorozotocin	-1.49	159.75	20	313.69	10	5	0	9	253.94
4	fumaric acid	-0.68	74.60	8	116.07	4	2	0	2	94.05
5	Geraniol	3.20	20.23	11	154.25	1	1	0	4	175.57
6	Glabridin	4.20	58.92	24	324.38	4	2	0	1	295.25
7	Glisoflavone	4.07	100.13	27	368.38	6	3	0	4	326.94
8	Licocoumarin	6.96	90.89	30	406.48	5	3	1	5	378.73
9	licoriphenone	4.44	96.22	27	372.42	6	3	0	7	343.72
10	Liquiritin	0.41	145.91	30	418.40	9	5	0	4	354.37
11	Isoangustone A	6.04	111.12	31	422.48	6	4	1	5	386.75

Molecular Docking:

Further docking is performed with the receptors in Table 7 with the above phytocompounds givne in Table 8-12. Docking scores, interacting amino acids along with number of interactions are noted in Table 13-25.

Table 13(a): Docking of MRE11 receptor with *Artemisia absinthium* :

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	p-cymene	-	-	NONE
2	Isoledene	-	-	NONE
3	Isopulegol acetate	-3802	PHE-327	1
4	Camphenol	-	-	NONE
5	Diisoamylene	-	-	NONE
6	Linalool	-3664	THR-153	1
7	n-amyl isovalerate	-3778	GLN-117	1

Table 13(b): Docking of MRE11 receptor with *Capparis spinosa*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	Apigenin	-4172	GLN-117,ASN-85	2
2	Flazin	-4762	ASP-156, SER-165, SER-165	3
3	Capparisin A	-	-	NONE
4	Kaempferol	-4306	ASP-156, ASN-85, GLN-117, VAL-119	4
5	Sakuranetin	-4384	SER-165,ASP-156	2
6	Stachydrine	-3000	ASN-85	1
7	Tetrahydroquinone	-2982	ALA-299	1
8	Thevetiaflavone	-	-	NONE

Table 13(c): Docking of MRE11 receptor with *Cleome drosiferila :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1-methylnapthalene	-	-	NONE
2	Acenaphthylene	-	-	NONE
3	Chlorpyrifos	-4634	GLN-117	1
4	Anthracene	-	-	NONE
5	Dichlorvos	-	-	NONE
6	Fenchlorphos	-	-	NONE

7	Lindane	-	-	NONE
8	Naphthalene	-	-	NONE

Table 13(d): Docking of MRE11 receptor with *Glycyrrhiza glabra :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1- methoxyficifolinol	-5876	ASN-85	1
2	Alpha terpineol	-3212	THR-153	1
3	Chlorozotocin	-4404	ASP-156, THR-153, GLN-117	3
4	fumaric acid	-2324	ASN-85, THR-153	2
5	Geraniol	-	-	NONE
6	Glabridin	-4980	ASN-85, THR-153	2
7	Glisoflavone	-5304	MSE-177,SER-115	2
8	Licocoumarin	-	-	NONE
9	Isoangustone A	-5840	SER-165, GLN-117	2
10	Licoriphenone	-5550	ASN-85, ASP-156	3
11	Liquiritin	-5708	SER-176, SER-203,ARG-175,MSE-177, ARG-208	5

Table 13(e): Docking of MRE11 receptor with *Harpagophytum procumbens:*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1,2-Benzenedicarboxylic acid,bis(2-methylpropyl) ester	-	-	NONE
2	6-Methoxynicotinic acid	-2902	LEU-158	2
3	2-(2,5-Dimethoxyphenyl)CYCLOHEX-2-ENONE	-	-	NONE
4	2,6-Dimethyl1-3-aminobenzoquinone	-3126	LYS-301, GLU-370, PHE-327	3
5	2-Methoxy-4-vinylphenol		-	NONE
6	Cyclohexanecarboxamide,N-furfuryl)	-4236	ASP-156	1
7	1S,2S,5R-1,4,4-Trimethyltricyclo[6.3.1.0(2,5)]dodec-8(9)-ene	-	-	NONE

Table 14(a): Docking of GSTP1 receptor with *Artemisia absinthium*:

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	p-cymene	-	-	NONE
2	Isoledene	-	-	NONE
3	Isopulegol acetate	-	-	NONE
4	Camphenol	-2760	ARG 14	1
5	Diisoamylene	-3376	GLN 126	2
6	Linalool	-	-	NONE
7	n-amyl isovalerate	-	-	NONE
8	2,4-heptadienal	-2946	ASP 99	1
9	Myrcene	-	-	NONE
10	Thymol	-3124	GLN 65 ,ASP-99	2

Table 14(b): Docking of GSTP1 receptor with *Capparis spinosa*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions

1	Apigenin	-3886	ARG14,ASP 95,ASN 67,GLN 65,ASO 95.	5
2	Flazin	-4454	GLU 98, ASP 99	2
3	Capparisin A	-3852	ASN 67, GLN 65,LEU53,GLN 65	4
4	Kaempferol	-4188	ARG 14, GLN 65	2
5	Sakuranetin	-4340	GLU 98,LEU 53, GLN 65	3
6	Stachydrine	-2664	ASP 99,ASN 67,GLU 98	3
7	Tetrahydroquinone	-2934	ASP 95	1
8	Thevetiaflavone	-4042	ASP 99 ,ASN 67	2

Table 14(c): Docking of GSTP1 receptor with *Cleome drosiferila :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1-methylnapthalene	-	-	NONE
2	Acenaphthylene	-	-	NONE
3	Chlorpyrifos	-4514	ASP 99	1
4	Anthracene	-	-	NONE
5	Dichlorvos	-3424	ASN 67 ,ASP 95	2
6	Fenchlorphos	-4364	ASN 67	1
7	Lindane	-	-	NONE
8	Naphthalene	-	-	NONE

Table 14(d): Docking of GSTP1 receptor with *Glycyrrhiza glabra :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1- methoxyficifolinol	-5544	GLN 52	1
2	Alpha terpineol	-3258	ASP 99	1
7	Chlorozotocin	-4178	GLN 52,GLN 65,ARG 14,LYS 103,GLU 98	5
4	fumaric acid	-2274	GLY 96,GLN 126,LYS 103	3
5	Geraniol	-3710	ASP 99	1
6	Glabridin	-4840	LEU 53,SER 66,ASN 67,ARG 14	4
7	Glisoflavone	-5238	GLN 52, ASP 99,ARG 14	3
8	Licocoumarin	-5658	LEU 53	1
9	Isoangustone A	-	-	NONE
10	Licoriphenone	-	-	NONE
11	Liquiritin	-5454	ARG 14,GLN 65,GLN 52,SER 66,GLN 65	5

Table 14(e): Docking of GSTP1 receptor with *Harpagophytum procumbens:*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1,2-Benzenedicarboxylic acid,bis(2-methylpropyl) ester	-4568	TYR 8, LEU 53	2
2	6-Methoxynicotinic acid	-3090	GLN 65,ASN 67	2
3	2-(2,5-Dimethoxyphenyl)CYCLOHEX-2-ENONE	-3758	ASP 99,GLN 65	2
4	2,6-Dimethyl1-3-aminobenzoquinone	-3054	ASP 95, GLU 98	2
5	2-Methoxy-4-vinylphenol	-3314	ARG 71, ASP 95	2
6	Cyclohexanecarboxamide,N-furfuryl)	-3758	ASP 99	1
7	1S,2S,5R-1,4,4-Trimethyltricyclo[6.3.1.0(2,5)]dodec-8(9)-ene	-	-	NONE

Table 15(a): Docking of CDKN2A receptor with *Artemisia absinthium:*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	p-cymene	-	-	None

2	Isoledene	-	-	None
3	Isopulegol acetate	-	-	None
4	Camphenol	-	-	None
5	Diisoamylene	-	-	None
6	Linalool	-	-	None
7	n-amyl isovalerate	-	-	None
8	2,4-heptadienal	-	-	None
9	Myrcene	-	-	None
10	Thymol	-	-	None

Table 15(b): Docking of CDKN2A receptor with *Capparis spinosa :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	Apigenin	-3216	GLU 23	1
2	Flazin	-3834	GLU 23	1
3	Capparisin A	-	-	NONE
4	Kaempferol	-3328	ARG 16	1
5	Sakuranetin	-3584	ARG 7	1
6	Stachydrine	-	-	NONE
7	Tetrahydroquinone	-	-	NONE
8	Thevetiaflavone	-	-	NONE

Table 15(c): Docking of CDKN2A receptor with *Cleome drosiferila*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1-methylnapthalene	-	-	None
2	Acenaphthylene	-	-	None
3	Chlorpyrifos	-	-	None
4	Anthracene	-	-	None
5	Dichlorvos	-3182	GLN 22	1
6	Fenchlorphos	-3496	ARG 7	1
7	Lindane	-	-	None
8	Naphthalene	-	-	None

Table 15(d): Docking of CDKN2A receptor with *Glycyrrhiza glabra*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1- methoxyficifolinol	-	-	None
2	Alpha terpineol	-	-	None
7	Chlorozotocin	-3338	GLN 22	1
4	fumaric acid	-1858	GLN 22	1
5	Geraniol	-2796	ARG 6	1
6	Glabridin	-	-	None
7	Glisoflavone	-	-	None
8	Licocoumarin	-4896	ARG 16 GLN 22	2
9	Isoangustone A	-4678	PHE 26	1
10	Licoriphenone	-	-	None
11	Liquiritin	-4370	ARG 7	1

Table 15(e): Docking of CDKN2A receptor with *Harpagophytum procumbens:*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1,2-Benzenedicarboxylic acid,bis(2-methylpropyl) ester	-4030	GLU 23 GLN 22	2
2	6-Methoxynicotinic acid	-	-	None
3	2-(2,5-Dimethoxyphenyl)CYCLOHEX-2-ENONE	-	-	None
4	2,6-Dimethyl1-3-aminobenzoquinone	-2306	ARG 6	1
5	2-Methoxy-4-	-	-	None

	vinylphenol			
6	Cyclohexanecarboxamide,N-furfuryl)	-	-	None
7	1S,2S,5R-1,4,4-Trimethyltricyclo[6.3.1.0(2,5)]dodec-8(9)-ene	-	-	None

Table 16(a): Docking of MIR21 receptor with *Artemisia absinthium:*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	p-cymene	-	-	None
2	Isoledene	-	-	None
3	Isopulegol acetate	-3860	SER 157, LEU 158,ARG 45	3
4	Camphenol	-3462	ASN 159	1
5	Diisoamylene	-4076	ASN 39, ASN 159	2
6	Linalool	-	-	None
7	n-amyl isovalerate	-4160	HIS 161,THR 38	2
8	2,4-heptadienal	-2874	ILE-149	1
9	Myrcene	-	-	None
10	Thymol	-3500	SER 157	1

Table 16(b): Docking of MIR21 receptor with *Capparis spinosa :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	Apigenin	-4400	THR 38,ASN 39,HIS 161	3
2	Flazin	-5006	THR 38,ASN 159,HIS 161	3
3	Capparisin A	-4088	ASN 39,HIS 70,ASN 159	3
4	Kaempferol	-4346	ARG 45,LYS 36,THR 38,ASN 159	4
5	Sakuranetin	-4614	ASN 159,ARG 45,ASN 159	3
6	Stachydrine	-3006	ASN 39,ASN 159	2
7	Tetrahydroquinone	-3028	LEU 158	1
8	Thevetiaflavone	-4646	LYS 36,HIS 70,THR 38,HIS 161	4

Table 16(c): Docking of MIR21 receptor with *Cleome drosiferila*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1-methylnapthalene	-	-	None
2	Acenaphthylene	-	-	None
3	Chlorpyrifos	-4604	HIS 70	1
4	Anthracene	-	-	None
5	Dichlorvos	-3542	ASN 159	1
6	Fenchlorphos	-4658	ASN 159	1
7	Lindane	-	-	None
8	Naphthalene	-	-	None

Table 16(d): Docking of MIR21 receptor with *Glycyrrhiza glabra*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1- methoxyficifolinol	-6438	ASN 39,THR 38	1
2	Alpha terpineol	-	-	None
7	Chlorozotocin	-4540	ARG 156,SER 157,LEU 158,GLU 131,GLY 129,GLU 107,VAL 106	7
4	fumaric acid	-2162	THR 38	1
5	Geraniol	-3720	ASN 159	1
6	Glabridin	-5442	ASN 159,LYS 36	2
7	Glisoflavone	-5296	HIS 70,HIS 161	2
8	Licocoumarin	-6080	SER 23	2
9	Isoangustone A	-6434	HIS 70, ASN 39	2
10	Licoriphenone	-5584	ASN 159,THR 38	2
11	Liquiritin	-6094	ASN 159 , THR 38,ASN 159	3

Table 16(e): Docking of MIR21 receptor with *Harpagophytum procumbens:*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1,2-Benzenedicarboxylic acid,bis(2-methylpropyl) ester	-4942	ASN 39, THR 38	2
2	6-Methoxynicotinic acid	-	-	None
3	2-(2,5-Dimethoxyphenyl)CYCLOHEX-2-ENONE	-4260	THR 38	1
4	2,6-Dimethyl1-3-aminobenzoquinone	-3060	ASN 159	1
5	2-Methoxy-4-vinylphenol	-3158	ASN 159	1
6	Cyclohexanecarboxamide,N-furfuryl)	-4326	ASN 39	1
7	1S,2S,5R-1,4,4-Trimethyltricyclo[6.3.1.0(2,5)]dodec-8(9)-ene	-	-	None

Table 17(a): Docking of POLQ receptor with *Artemisia absinthium:*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	p-cymene	-	-	None
2	Isoledene	-	-	None
3	Isopulegol acetate	-3966	ARG 639	1
4	Camphenol	-	-	None
5	Diisoamylene	-3768	HIS 651	1
6	Linalool	-4092	ARG 639,ASP 636,ARG 713	3
7	n-amyl isovalerate	-4114	ASP 636	1
8	2,4-heptadienal	-	-	None
9	Myrcene	-	-	None
10	Thymol	-	-	None

Table 17(b): Docking of POLQ receptor with *Capparis spinosa :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	Apigenin	-4294	HIS 651,ARG 639,ARG 713	3
2	Flazin	-4842	ASP 636,HIS 651	2
3	Capparisin A	-4018	ARG 713,ARG 639	2
4	Kaempferol	-4330	TYR 654,HIS 651,ARG 639	3
5	Sakuranetin	-4648	ARG 713,HIS 651	2
6	Stachydrine	-	-	None
7	Tetrahydroquinone	-	-	None
8	Thevetiaflavone	-4678	VAL 689,HIS 651	2

Table 17(c): Docking of POLQ receptor with *Cleome drosiferila :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1-methylnapthalene	-	-	None
2	Acenaphthylene	-	-	None
3	Chlorpyrifos	-4872	ARG 202	1
4	Anthracene	-	-	None
5	Dichlorvos	-3484	ARG 808	1
6	Fenchlorphos	-4602	ARG 639	1
7	Lindane	-	-	None
8	Naphthalene	-	-	None

Table 17(d): Docking of POLQ receptor with *Glycyrrhiza glabra :*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions

1	1- methoxyficifolinol	-6310	ARG 195	1
2	Alpha terpineol	-	-	None
7	Chlorozotocin	-4272	TYR 654,VAL 689,HIS 651,ASP 636,ASP 639,ALA 640	6
4	fumaric acid	-2260	TYR 654,LEU 688,ARG 639	3
5	Geraniol	-	-	None
6	Glabridin	-4906	ARG 639,HIS 651,TYR 654	3
7	Glisoflavone	-5306	THR 867,GLN 751	2
8	Licocoumarin	-5906	THR 119	1
9	Isoangustone A	-6056	ASN 805,ALA 808,GLN 754	3
10	Licoriphenone	-5554	ASP 636,TYR 654,ARG 713	3
11	Liquiritin	-5646	ASP 632,HIS 651	2

Table 17(e): Docking of POLQ receptor with *Harpagophytum procumbens*

S.no	Compound	Docking score	Interacting amino acid	No.of Interactions
1	1,2-Benzenedicarboxylic acid,bis(2-methylpropyl) ester	-	-	None
2	6-Methoxynicotinic acid	-3054	ARG 398	1
3	2-(2,5-Dimethoxyphenyl)CYCLOHEX-2-ENONE	-4164	VAL 689	1
4	2,6-Dimethyl1-3-aminobenzoquinone	-3126	THR 863,TRP 865	2
5	2-Methoxy-4-vinylphenol	-	-	None
6	Cyclohexanecarboxamide,N-furfuryl)	-4288	ASP 636,VAL 689	2
7	1S,2S,5R-1,4,4-Trimethyltricyclo[6.3.1.0(2,5)]dodec-8(9)-ene	-	-	None

As per the docking results it is seen that the compounds Apigenin, Chlorozotocin and 2,6-Dimethyl1-3-aminobenzoquinone docks with good interactions with all the receptors mentioned in table and it shows no violations in Lipinski's rule of five.

As per docking studies it is seen that MRE11 receptor docks with Apigenin with a docking score of -4172kcal/mol and interacts with the amino acids GLN-117 and ASN-85, GSTP1 receptor docks with a docking score of -3886kcal/mol and interacts with the amino acids ARG14,ASP 95,ASN 67,GLN 65 and ASO 95, CDKN2A receptor docks with a docking score of -3216 kcal/mol and interacts with the amino acid GLU 23, MIR21 receptor docks with a docking score of -4400kcal/mol and interacts with the amino acids THR 38,ASN 39 and HIS 161 and POLQ receptor docks with a docking score of -4294kcal/mol and interacts with the amino acids HIS 651,ARG 639 and ARG 713.

Again, as per docking studies it is seen that MRE11 receptor docks with Chlorozotocin with a docking score of -4404kcal/mol and interacts with the amino acids ASP-156, THR-153 and GLN-117, GSTP1 receptor docks with a docking score of -4178kcal/mol and interacts with the amino acids GLN 52,GLN 65,ARG 14,LYS 103 and GLU 98, CDKN2A receptor docks with a docking score of -3338kcal/mol and interacts with the amino acid GLN 22, MIR21 receptor docks with a docking score of -4540kcal/mol and interacts with the amino acids ARG 156,SER 157,LEU 158,GLU 131,GLY 129,GLU 107 and VAL 106

and POLQ receptor docks with a docking score of -4272kcal/mol and interacts with the amino acids TYR 654,VAL 689,HIS 651,ASP 636,ASP 639 and ALA 640.

Further, as per docking studies it is seen that MRE11 receptor docks with2,6-Dimethyl1-3-aminobenzoquinone with a docking score of -3126kcal/mol and interacts with the amino acids LYS-301, GLU-370 and PHE-327, GSTP1 receptor docks with a docking score of -3054kcal/mol and interacts with the amino acids ASP 95 and GLU 98, CDKN2A receptor docks with a docking score of -2306kcal/mol and interacts with the amino acid ARG 6, MIR21 receptor docks with a docking score of-3060kcal/mol and interacts with the amino acid ASN 159 and POLQ receptor docks with a docking score of -3126 kcal/mol and interacts with the amino acids THR 863 andTRP 865.

CONCLUSION

The taxonomy and gene family information of human oropharynx microbiome are identified. Again, as per docking studies and ADME analysis it is seen that phytocompounds Apigenin, Chlorozotocin and 2,6-Dimethyl1-3-aminobenzoquinone can be potential ligands for the receptors implicated in oropharynx. Further, *in-vitro* and *in-vivo* studies can be done on the above phytocompounds to establish their potential as drugs in treating oropharynx.

REFERENCE

1. Fujiwara N, Kitamura N,Yoshida K, Yamamoto T, Ozaki K, and Kudo Y, 2020, Involvement of Fusobacterium Species in Oral Cancer Progression: A Literature Review Including Other Types of Cancer, Int J Mol Sci. 21(17): 6207.

2. Zhang L, Liu Y, Zheng HJ and Zhang CP, 2020, The Oral Microbiota May Have Influence on Oral Cancer, Front. Cell. Infect. Microbiol. 9:476.

3. Miranda-Galvis M, Loveless R, Kowalski L.P, Teng Y, Impacts of Environmental Factors on Head and Neck Cancer Pathogenesis and Progression. Cells 2021, 10, 389.

4. Abed J, Maalouf N, Manson AL, Earl AM, Parhi L, Emgård JEM, Klutstein M, Tayeb S, Almogy G, Atlan KA, Chaushu S, Israeli E, Mandelboim O, Garrett WS and Bachrach G (2020) Colon Cancer-Associated Fusobacterium nucleatum May Originate From the Oral Cavity and Reach Colon Tumors via the Circulatory System. Front. Cell. Infect. Microbiol. 10:400.

5. Olsen I. (2014) The Family Fusobacteriaceae. In: Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F. (eds) The Prokaryotes. Springer, Berlin, Heidelberg.

6. Broadley, Marissa MPH, MSEd, RN, CIC; Schweon, Steven J. MSN, MPH, RN, CIC, HEM, FSHEA, FAPIC Get the facts about Fusobacterium, Nursing: 2017, 47(5):64-65.

7. Brennan CA, Garrett WS. Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. Nat Rev Microbiol. 2019;17(3):156-166.

8. Candela T, Moya M, Haustant M, Fouet A. Fusobacterium nucleatum, the first Gram-negative bacterium demonstrated to produce polyglutamate. Can J Microbiol. 2009, 55(5):627-32.

9. Finegold SM. Anaerobic Gram-Negative Bacilli. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 20. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8438/

10. Deo PN, Deshmukh R. Oral microbiome: Unveiling the fundamentals. J Oral Maxillofac Pathol. 2019;23(1):122-128. doi:10.4103/jomfp.JOMFP_304_18

11. Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol. 2015;23:141-147. doi:10.1016/j.mib.2014.11.013

12. Li X, Kolltveit KM, Tronstad L, Olsen I. Systemic diseases caused by oral infection. Clin Microbiol Rev. 2000;13(4):547-558.

13. Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Microb Inform Exp. 2012;2(1):3. Published 2012 Feb 9. doi:10.1186/2042-5783-2-3

14. Alejandra EZ, de León Arturo VP, Alejandro SF, 2015, The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics, Frontiers in Genetics, 6:348

15. Sharpton Thomas J, 2015, An introduction to the analysis of shotgun metagenomic data, Frontiers in Plant Science, 5, 209.

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