Establishing the Taxa with Phylogenetic Profile and *in-silico-ayurvedic* Remedy of Colon Cancer Microbiome

Sai Divya Paggala[1(✉)], I. A. Shylesh Murthy[2], and Preenon Bagchi[1,2,3]

[1] Padmashree Institute of Management and Science, Bangalore, India
saidivya.p13@gmail.com
[2] Vasishth Academy of Advanced Studies and Research (Sarvasumana Association), Bangalore, India
[3] MGM Institute of Biosciences and Technology, Aurangabad, India

Abstract. The discovery of a bacterium, *Helicobacter pylori* that is resident in the human stomach chronic disease (peptic ulcer and gastric cancer) was radical on many levels. Studies of this genetic diversity in strains isolated from various locations across the globe. The goal of this study was to determine the role of 16S ribosomal RNA (rRNA) was used to compare the microbiota into new angles to dig out potential in metagenomics and molecular docking studies of colon cancer.

Keywords: Chronic Disease · Genetic Diversity · rRNA · Microbiota · Metagenomics · Molecular docking · Colon cancer

1 Introduction

Helicobacter pylori (*H. pylori*) is a gram-negative microaerophilic bacterium that colonizes the gastric mucosa of more than half of the worldwide population with high geographic variability [1, 2]. *H. pylori* infection is generally acquired during childhood and can persist life-long without symptoms. It triggers pathogenesis by creating reactive oxygen species and modulating host-inflammatory responses. This pathogen is known to cause diseases of the upper gastrointestinal tract such as peptic ulcer, gastric cancer, and gastric mucosa-associated lymphoid tissue (MALT) lymphoma [3]. In addition, host genetic, immunogenic factors, and environmental factors including resident gut microbiota is known to play a significant role in disease pathogenesis [4].

Pathogenicity

Colorectal cancer is the second- and third-most common cancer in women and men, respectively. In 2012, 614,000 women (9.2% of all new cancer cases) and 746,000 men (10.0% of new cancer cases) were diagnosed with colorectal cancer worldwide [5]. Combined, in both sexes, colorectal cancer is the third-most common cancer and accounts for 9.7% of all cancers excluding non-melanoma skin cancer. More than half of the cases occur in more-developed regions of world. The age-standardized incidence rate (ASRi) of colorectal cancer is higher in men (20.6 per 100,000 individuals) than in

R. Somashekhar et al. (Eds.): ICBDS 2022, AHSR 58, pp. 159–177, 2023.
https://doi.org/10.2991/978-94-6463-164-7_13

women (14.3 per 100,000). The majority of patients with sporadic cancer are >50 years of age, with 75% of patients with rectal cancer and 80% of patients with colon cancer patients being ≥60 years of age at the time of diagnosis [6]. Colorectal cancer accounts for approximately 10% of all annually diagnosed cancers and cancer-related deaths worldwide. Both hereditary and environmental risk factors play a part in the development of colorectal cancer. Hereditary colorectal cancer syndromes can be subdivided as non-polyposis (Lynch syndrome and familial colorectal cancer) and polyposis syndromes [7]. Both genetic and environmental factors play an important part in the aetiology of colorectal cancer. The majority of colorectal cancers are sporadic; approximately three-quarters of patients have a negative family history [8]. In most Western populations, the average lifetime risk for colorectal cancer is in the range of 3–5%. However, this risk almost doubles in individuals with a first-degree family member with colorectal cancer who was diagnosed at 50–70 years of age; the risk triples if the first-degree relative was <50 years of age at diagnosis [9]. Risk further increases in individuals who have two or more affected family members. For sporadic colorectal cancer, this increased risk in the presence of affected family at least in part reflects low-penetrance genetic factors. Accordingly, positive family history has a role in approximately 15–20% of patients with colorectal cancer [10]. Indeed, a specific subgroup of the patient population is formed by those affected by a hereditary colorectal cancer syndrome, accounting for 5–10% of all patients. The most common syndrome in this category is Lynch syndrome [11]. This syndrome is caused by a mutation in one of the DNA mismatch-repair genes: MLH1, MSH2, MSH6, PMS2 or EPCAM. Impaired mismatch repair during replication gives rise to accumulation of DNA mutations, which occur, in particular, in microsatellite DNA fragments with repetitive nucleotide sequence. This microsatellite instability (MSI) can be identified by means of polymerase chain reaction (PCR) testing, which compares normal and tumor DNA of the same patient [12]. Patients with Lynch syndrome used to be identified by means of clinic pathological criteria, such as the Amsterdam and Bethesda criteria. However, clinical practice is shifting towards unrestricted testing of tumor material of all patients diagnosed before the age of 70 years by means of MSI PCR and immunohistochemistry for lack of expression of specific mismatch-repair proteins [13].

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) [13].

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 [13]. 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 [14–16]. 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 16S Microbial Analysis with mothur by Hiltemann S, Batut B and Clements D is used to analyze colon cancer microbiome [17]. *Helicobacter pylori* (*H. pylori*)' fastq sequences DRR286990, SRR13787057 and SRR1416071 were retrieved from SRA database. Sequences were paired to make collection of paired datasets. Contigs were created from paired-end reads using Make.contigs tool. Next, we take a summary of our data using Summary.seqs tool. Further, we improve the quality of our data and data cleaning using the Screen.seqs tool, which removes sequences with ambiguous bases and contigs longer than a given threshold. Next, we will remove any overhang sequence using Filter.seqs tool. Next, in order to speed up computation work, we determine the unique reads Unique.seqs tool and then record how many times each of these different reads was observed in the original dataset using Count.seqs tool. Next we do Sequence Alignment using Align.seqs tool and get a summary of our sequences using Summary.seqs tool. Next we remove any non-overlapping reads using Screen.seqs tool, remove any overhang on either end to ensure our sequences overlap using Filter.seqs tool. Next, we deduplicate our data by re-running the Unique.seqs tool. Further, we perform preliminary clustering of sequences with Pre.cluster tool. Further, we classify the sequences into phylotypes using Classify.seqs [18]. 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 [19]. Further, we see how many sequences we have in each sample with count.groups tool and do subsampling using Sub.sample tool.

Next, we generate Venn diagrams using Venn tool. Further, we perform beta diversity using Dist.shared tool. Further, we generated the dendrogram to describe the similarity of our sequences with each other. We generated dendrogram using the jclass and thetayc calculators within the Tree.shared tool.

Next, using the genes present in colon cancer microbiome, their 3d structure was modeled using SWISS-MODEL [20]. Phyto-compounds were downloaded from PUBCHEM. Using, molinspiration software [21], following the principles of Lipinski's rule of five, phytocompounds were selected for docking. Further, docking was performed using patchdock [22].

3 Results and Discussion

Helicobacter pylori (*H. pylori*)' combined fastq sequences DRR286990, SRR13787057 and SRR1416071 data summary as per summary.seq is given in Table 1.

Table 1. Summary.seq output of combined fastq sequences DRR286990, SRR13787057 and SRR1416071

Start	End	NBases	Ambigs	Polymer	NumSeqs	
Minimum:	1	128	128	0	2	1
2.5%-tile:	1	173	173	0	4	659330
25%-tile:	1	239	239	1	5	6593291
Median:	1	268	268	6	6	13186581
75%-tile:	1	286	286	16	6	19779871
97.5%-tile:	1	299	299	40	8	25713832
Maximum:	1	350	350	85	150	26373160
Mean:	1	258.928	258.928	10.2347	5.71981	
# of Seqs:	26373160					

Table 2. Align.seq output (1st few lines)

QueryName	Query Length	Template Name	Template Length	Search Method	Search Score	Alignment Method	Pairwise Alignment Length
14_DRR286990	270	AJ009481.1	293	kmer	2.66	needleman	2
28_DRR286990	270	AF393378.1	293	kmer	2.66	needleman	10
39_DRR286990	334	AF445690.1	295	kmer	2.45	needleman	2
45_DRR286990	346	AJ290045.1	293	kmer	3.24	needleman	4
46_DRR286990	270	AF139200.1	293	kmer	3.42	needleman	3
50_DRR286990	264	AF352532.1	294	kmer	3.11	needleman	-558
55_DRR286990	346	AF445705.1	293	kmer	2.65	needleman	5
61_DRR286990	338	X70907.1	292	kmer	3.32	needleman	14
104_DRR286990	235	AB015546.1	293	kmer	2.63	needleman	3
144_DRR286990	268	AY493581.1	293	kmer	3.83	needleman	30
145_DRR286990	346	U42638.1	292	kmer	2.65	needleman	3
194_DRR286990	272	AY532580.1	292	kmer	2.64	needleman	7
218_DRR286990	264	AF445690.1	295	kmer	3.50	needleman	1
222_DRR286990	266	U81676.1	293	kmer	3.09	needleman	10
241_DRR286990	348	AF416764.1	293	kmer	2.35	needleman	5
260_DRR286990	272	AF449770.1	293	kmer	3.02	needleman	15
262_DRR286990	264	AJ231168.1	293	kmer	3.11	needleman	6

(*continued*)

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. First few lines of the Align.seq output is given in Table 2.

Table 2. (*continued*)

QueryName	Query Length	Template Name	Template Length	Search Method	Search Score	Alignment Method	Pairwise Alignment Length
263_DRR286990	344	AJ241002.1	292	kmer	2.37	needleman	2
273_DRR286990	332	AF050596.1	294	kmer	2.15	needleman	5
280_DRR286990	248	AY033322.1	293	kmer	2.90	needleman	6
282_DRR286990	268	AY218548.1	294	kmer	3.07	needleman	6
297_DRR286990	344	AY345577.1	291	kmer	2.97	needleman	5
329_DRR286990	348	EF126993.1	293	kmer	2.35	needleman	3
332_DRR286990	264	AY212681.1	295	kmer	3.50	needleman	9
359_DRR286990	344	AY492086.1	292	kmer	2.67	needleman	2
365_DRR286990	348	U22013.1	295	kmer	2.64	needleman	1
385_DRR286990	348	AF449770.1	293	kmer	3.81	needleman	5
388_DRR286990	270	AF269001.1	294	kmer	3.04	needleman	5
390_DRR286990	268	AJ224942.1	291	kmer	3.07	needleman	1
400_DRR286990	264	AY214200.1	293	kmer	2.72	needleman	1
421_DRR286990	342	AF280847.1	296	kmer	1.79	needleman	4
495_DRR286990	344	AJ344553.1	293	kmer	2.37	needleman	9
503_DRR286990	346	DQ501304.1	292	kmer	2.06	needleman	9
512_DRR286990	266	AY328606.1	293	kmer	3.47	needleman	5
519_DRR286990	308	AF094732.1	294	kmer	2.99	needleman	4
526_DRR286990	268	AB064919.1	293	kmer	3.83	needleman	9
528_DRR286990	270	AJ290033.1	293	kmer	3.42	needleman	7

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

Next, we removed any overhang on either side of the V4 region using Filter.seqs tool. Tool's output summary is given in below table.

```
Length of filtered alignment: 15
Number of columns removed: 13410
Length of the original alignment: 13425
Number of sequences used to construct filter: 463776
```

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

Table 3. Summary.seq output of Align.seq

	Start	End	NBases	Ambigs	Polymer	NumSeqs	
Minimum:		0	0	0	0	1	1
2.5%-tile:		0	0	0	0	1	133964
25%-tile:		1	1248	2	0	1	1339636
Median:		13399	13425	4	0	2	2679271
75%-tile:		13422	13425	8	0	2	4018906
97.5%-tile:		13425	13425	24	0	4	5224578
Maximum:		13425	13425	297	0	31	5358541
Mean:		9209.21	9624.39	6.01429	0	1.8588	
# of unique seqs:			3692641				
total # of seqs:			5358541				

Table 4. Taxonomy output of Classify.seq

taxlevel	**rankID**	**taxon**	**daughterlevels**	**total**
Taxonomy	total	DRR286990	SRR13787057	T5CM_L8
Root	677007	14336	132613	530058
unknown;unknown_unclassified; unknown_unclassified; unknown_unclassified;unknown_unclassified; unknown_unclassified;	677007	14336	132613	530058

Count.group gave the count of the sequences as given in below table

DRR286990 contains 14336.
SRR13787057 contains 132613.
T5CM_L8 contains 530058.
Total seqs: 677007.

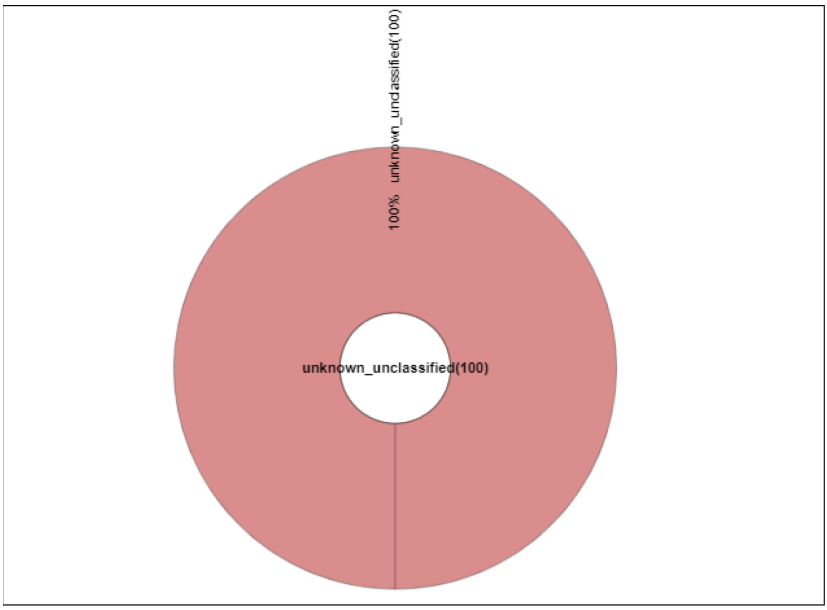

Fig. 1. Krona pie chart visualization of the taxonomy

Venn Diagram of the sequences gave the output

Structure Based Drug Designing of Colon Cancer

Since, colon cancer is a disease caused by bacteria, we further go ahead towards designing novel drug for the disease. The gene receptors corresponding to colon cancer are taken from NCBI for our work (Table 5).

Abbreviations of Genes:

1) MSH6: MutS Homolog 6
2) MSH2: MutS Homolog 2
3) MLH1: MutL Homolog 1

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 5.

Ayurvedic Medicinal plants *Ocimum sanctum*, *Cinnamomum cassia*, C*yperus rotundus*, *Brassica juncea* and *Sonchus arvensis* many bacterial and cancer related diseases [23]. The potency of their phytocompounds in treating Colon Cancer 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 (Tables 6, 7, 8, 9 and 10).

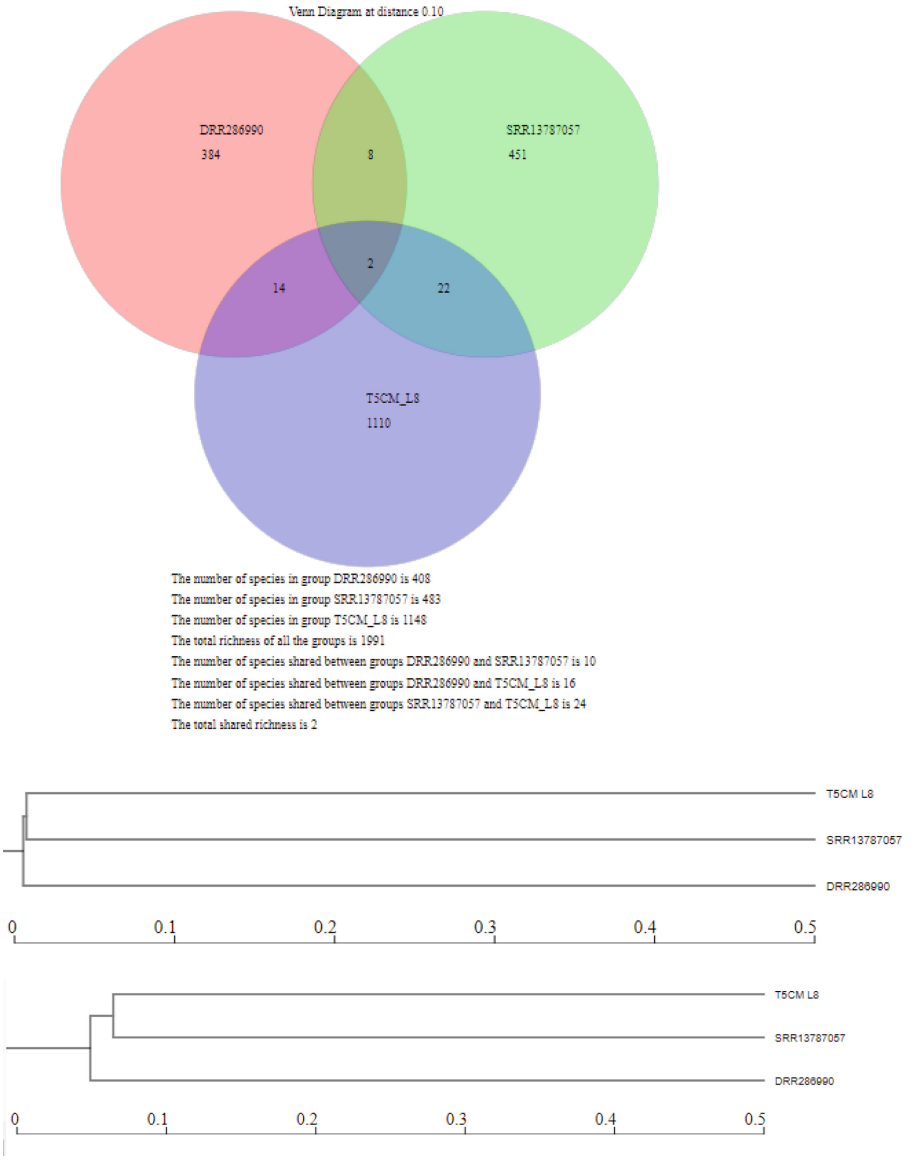

Fig. 2. Venn diagram (top) and phylogenetic tree (below) visualization of the taxonomy

Molecular Docking

Further docking is performed with the receptors in Table 7 with the above phytocompounds. Docking scores, interacting amino acids along with number of interactions are noted in Tables 11, 12, 13, 14 and 15.

As per the docking results it is seen that MLH1 receptor docks with alphaethylidene benzeneacetaldehyde with a docking score of 2752 kcal/mol and with 1 interaction,

Table 5. Genes with their NCBI Accession number

Sl. No	Gene Receptors	NCBI Accession Number	Homologous Template
1.	MSH6	P52701.2	208BD
2.	MSH2	P43246.1	3THWA
3.	MLH1	ACR33810.1	3RBNA

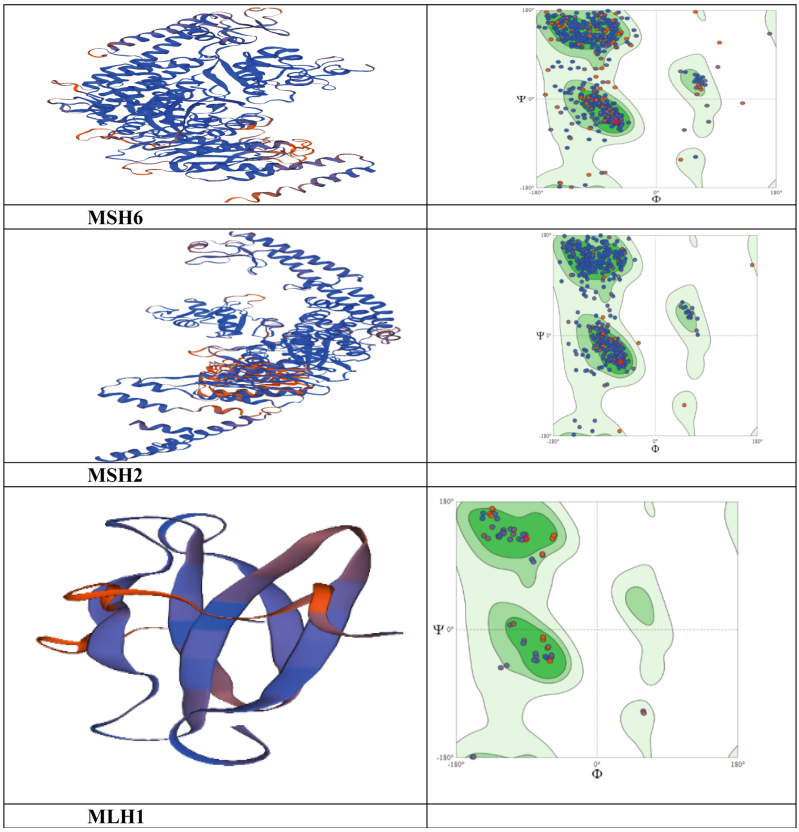

Fig. 3. Swiss-model generated receptor models with their ramachandran plot

betacyclocitral with a docking score of 2554 kcal/mol with 1 interaction, cyclohexanol with a docking score of 1956 kcal/mol with 1 interaction, Dibutyl phthalate with a docking score of 3878 kcal/mol with 1 interaction and Pyrrol-2-one with a docking score of 1940 kcal/mol with 1 interaction. MSH2 docks with alphaethylidenebenzeneacetaldehyde with a docking score of 3268 kcal/mol with 1 interaction, benzeneacetaldehyde with a docking score of 3008 kcal/mol with 2 interactions, betacyclocitral with a docking score of 3048 kcal/mol with 1 interaction, Betaionone with a docking score of 3748 kcal/mol with 2 interactions, cyclohexanol with a docking score of 2628 kcal/mol

Table 6. ADME studies of *Ocimum santanum*

Sno.	Compound	Mi logP	TPS	natoms	mw	nON	nOHNH	Nviolation	nrtob	volume
1	cinnamaldehyde	2.48	17.07	10	132.16	1	0	0	2	130.44
2	Diacetone alcohol	0.31	37.30	8	116.16	2	1	0	2	122.62
3	Estragole	2.82	9.23	11	148.21	1	0	0	3	154.12
4	Eucalyptol	2.72	9.23	11	154.25	1	0	0	0	166.66
5	Eugenol	2.10	29.46	12	164.20	2	1	0	3	162.14
6	Linalool	3.21	20.23	11	154.25	1	1	0	4	175.59
7	Methyleugenol	2.48	18.47	13	178.23	2	0	0	4	179.67

Table 7. ADME studies of *Cinnamomum cassia*

Sno	Compound	Mi logP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrtob	volume
1	Acetophenone	1.84	17.07	9	120.15	1	0	0	1	119.59
2	Alphabisabolol	4.68	20.23	16	222.37	1	1	0	4	248.23
3	Alphaterpineol	2.60	20.23	11	154.25	1	1	0	1	170.65
4	Alpha-thujene	3.31	0.00	10	136.24	0	0	0	1	151.81
5	Borneol	2.35	20.23	11	154.25	1	1	0	0	165.72
6	Cis-cinnamaldehyde	2.48	17.07	10	132.16	1	0	0	2	130.44
7	Linalool	3.21	20.23	11	154.25	1	1	0	4	175.59
8	Styrene	2.79	0.00	8	104.15	0	0	0	1	111.78

Table 8. ADME studies of *Cyperus rotundus*

Sno.	Compounds	MilogP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	volume
1	Mytenol	2.30	20.23	11	152.24	1	1	0	1	160.07
2	Nookatone	3.67	17.07	16	218.34	1	0	0	1	232.13
3	2-propenoic acid	0.20	37.30	5	72.06	2	1	0	1	67.37
4	Sabinene	3.10	0.00	10	136.24	0	0	0	1	152.37
5	Cryprotene	4.57	0.00	14	192.35	0	0	0	0	218.24
6	Verbenone	2.44	17.07	11	150.22	1	0	0	0	154.00
7	Carvone	2.51	17.07	11	150.22	1	0	0	1	159.48
8	Caryophyllene epoxide	4.14	12.53	16	220.36	1	0	0	0	234.01

Table 9. ADME studies of *Brassica juncea*

Sno.	compounds	MilogP	TPSA	natoms	mw	nON	nOHNH	nviolatons	nROTB	volume
1	alpha-ethylidene-benzeneacetaldehyde	2.36	17.07	11	146.19	1	0	0	2	147.00
2	benzeneacetaldehyde	1.94	17.07	9	120.15	1	0	0	2	119.83
3	beta-Cyclocitral	3.20	17.07	11	152.24	1	0	0	1	164.78
4	beta-ionone	3.45	17.07	14	192.30	1	0	0	2	208.76
5	Cyclohexanol	1.59	20.23	7	100.16	1	1	0	0	110.65
6	Dibutyl phthalate	4.43	52.61	20	278.35	4	0	0	10	273.91
7	Dimethyl trisulfide	1.84	0.00	5	126.27	0	0	0	2	100.14
8	pyrrol-2-one	0.76	36.02	6	83.09	2	2	0	0	77.05

Table 10. ADME studies of *Sonchus arvensis*

Sno.	compounds	MilogP	TPSA	natoms	mw	nON	nOHNH	nviolatons	nROTB	volume
1	3-(2-Methoxyethyl)-1-nonanol	4.99	18.47	18	278.50	2	0	0	12	302.33
2	3-Methylnonane	4.89	0.00	10	142.29	0	0	0	6	179.96
3	Bornyl acetate	3.05	26.30	14	196.29	2	0	0	2	202.23
4	Camphor	2.16	17.07	11	152.24	1	0	0	0	159.86
5	Isobutyl phthalate	3.80	52.61	20	278.35	4	0	0	8	273.48
6	p-Xylene	2.83	0.00	8	106.17	0	0	0	0	117.17
7	Toluene	2.39	0.00	7	92.14	0	0	0	0	100.60
8	trans-1,3-Dimethylcyclopentane	2.48	0.00	7	98.19	0	0	0	0	118.98

with 1 interaction, Dibutyl phthalate with a docking score of 4470 kcal/mol with 1 interaction, Pyrrol-2-one with a docking score of 2216 kcal/mol with 1 interaction. MSH6 docks with alphaethylidenebenzeneacetaldehyde with a docking score of 3326 kcal/mol with 1 interaction, benzeneacetaldehyde with a docking score of 2782 kcal/mol with 1 interaction, betacyclocitral with a docking score of 2952 kcal/mol with 1 interaction, Betaionone with a docking score of 3926 kcal/mol with 1 interaction, cyclohexanol with a docking score of 2424 kcal/mol with 1 interaction, Dibutyl phthalate with a docking score of 5052 kcal/mol with 2 interactions, Pyrrol-2-one with a docking score of 1972 kcal/mol with 1 interaction.

Again, it is seen that MLH1 docks with Diacetonealcohol with a docking score of 2450 kcal/mol with 1 interaction. Also, it is seen that MSH6 docks with Methyleugenol with a docking score of 3882 kcal/mol with 1 interaction, Eugenol with a docking score of 3538 kcal/mol with 1 interaction, cinnamaldehye with a docking score of 3008 kcal/mol with 1 interaction, Estragole with a docking score of 3390 kcal/mol with 1 interaction, Diacetonealcohol with a docking score of 2674 kcal/mol with 1 interaction, Linalool with a docking score of 3634 kcal/mol with 1 interaction. Further, it is seen that MSH2 docks with Cinnamaldehyde with a docking score of 3270 kcal/mol with 1 interaction, Diacetonealcohol with a docking score of 2906 kcal/mol with 2 interactions, Estragole with a docking score of 3466 kcal/mol with 1 interaction, Linalool with a docking score of 3740 kcal/mol with 2 interactions, Eucalyptol with a docking score of 3048 kcal/mol with

Table 11. Docking of receptors with *Brassica juncea*

Sl. No.	Receptor protein	Compound	Docking score (-kcal/mol)	Interacted amino acid	No. of interactions
1	MLH1	Alphaethylidene benzeneacetaldehyde	2752	LEU -1	1
2	MLH1	benzeneacetaldehyde	-	-	-
3	MLH1	Betacyclocitral	2554	THR-31	1
4	MLH1	Betaionone	-	-	-
5	MLH1	cyclohexanol	1956	HIS-6	1
6	MLH1	Dibutyl phthalate	3878	HIS-8	1
7	MLH1	Dimethyltrisulfide	-	-	-
8	MLH1	Pyrrol-2-one	1940	LEU-1	1

Sl. No.	Receptor protein	Compound	Docking score	Interacted amino acid	No.of interactions
1	MSH2	alphaethylidenebenzeneacetaldehyde	3268	GLY-220	1
2	MSH2	benzeneacetaldehyde	3008	SER-129 GLY-130	2
3	MSH2	betacyclocitral	3048	HIS-785	1
4	MSH2	Betaionone	3748	GLU-278 LEU-277	2
5	MSH2	cyclohexanol	2628	PHE-136	1
6	MSH2	Dibutyl phthalate	4470	GLY-220	1
7	MSH2	Dimethyltrisulfide	-	-	-
8	MSH2	Pyrrol-2-one	2216	PHE-136	1

Sl. No.	Receptor Protein	Compound	Docking Score	Interacted amino acid	No. of interaction
1	MSH6	Alphaethylidene benzeneacetaldehyde	3326	ASP-838	1
2	MSH6	benzeneacetaldehyde	2782	ASP-838	1
3	MSH6	betacyclocitral	2952	GLU-618	1
4	MSH6	Betaionone	3926	VAL-827	1
5	MSH6	cyclohexanol	2424	SER-393	1
6	MSH6	Dibutyl phthalate	5052	LYS-832 ASP-838	2
7	MSH6	Dimethyltrisulfide	-	-	-
8	MSH6	Pyrrol-2-one	1972	TYR-1286	1

Table 12. Docking of receptors with *Ocimum santanum*

Sl. no	Receptor protein	Compound	Docking score (-kcal/mol)	Interacted amino acid	No. of interactions
1	MLH1	Cinnamaldehyde	-	-	-
2	MLH1	Diacetonealcohol	2450	LEU-1	1
3	MLH1	Estragole	-	-	-
4	MLH1	Eucalyptol	-	-	-
5	MLH1	Eugenol	-	-	-
6	MLH1	Linalool	-	-	-
7	MLH1	methyleugenol	-	-	-

Sl. no	Receptor Protein	Compound	Docking score	Interacted aminoacid	No.of interactions
1	MSH6	Methyleugenol	3882	LYS-853	1
2	MSH6	Eugenol	3538	GLN-625	1
3	MSH6	cinnamaldehye	3008	ASP-838	1
4	MSH6	Eucalyptol	-	-	-
5	MSH6	Estragole	3390	LYS-853	1
6	MSH6	Diacetonealcohol	2674	GLU-618	1
7	MSH6	Linalool	3634	LYS-832	1

Sl. no	Receptor Protein	Compound	Docking score	Interacted aminoacid	No.of interactions
1	MSH2	Cinnamaldehyde	3270	GLU-278	1
2	MSH2	Diacetonealcohol	2906	ASP-282, LYS-392	2
3	MSH2	Estragole.	3466	ASN-285	1
4	MSH2	Linalool	3740	LYS-392, SER-129	2
5	MSH2	Eucalyptol	3048	SER-129	1
6	MSH2	Methyleugenol	3592	GLN-183	1
7	MSH2	Eugenol	3598	GLN-130, SER-284	2

1 interaction, Methyleugenol with a docking score of 3592 kcal/mol with 1 interaction and Eugenol with a docking score of 3598 kcal/mol with 2 interactions.

Again, it is seen that MLH1 docks with Acetophenone with a docking score of 2578 kcal/mol with 1 interaction, Borneol with a docking score of 2318 kcal/mol with 1 interaction and Cis-cinnamaldehyde with a docking score of 2336 kcal/mol with 1 interaction. Further, it is seen that MSH2 docks with Acetophenone with a docking score of 3046 kcal/mol with 1 interaction, Alpha-bisabolol with a docking score of

Table 13. Docking of receptors with *Cinnamomum cassia*

Sl. no	Receptor Protein	Compound	Docking score score (-kcal/mol)	Interacted amino acid	No.of interactions
1	MLH1	Acetophenone	2578	HIS - 6	1
2	MLH1	Alpha-bisabolol	-	-	-
3	MLH1	Alpha-terpineol	-	-	-
4	MLH1	Alpha-thujene	-	-	-
5	MLH1	Borneol	2318	HIS -8	1
6	MLH1	Cis-cinnamaldehyde	2336	LEU -1	1
7	MLH1	Linalool	-	-	-
8	MLH1	Styrene	-	-	-

Sl. no	Receptor Protein	Compound	Docking score	Interacted amino acid	No.of interactions
1	MSH2	Acetophenone	3046	LEU-277	1
2	MSH2	Alpha-bisabolol	4372	LEU-277	1
3	MSH2	Alpha-terpineol	3380	ASN -285, ARG-171	2
4	MSH2	Alpha-thujene	-	-	-
5	MSH2	Borneol	2880	HIS-783	1
6	MSH2	Cis-cinnamaldehyde	3190	ASP-282, GLN -130	2
7	MSH2	Linalool	3740	SER-129, LYS-392	2
8	MSH2	Styrene	-	-	-

Sl. no	Receptor Protein	Compound	Docking score	Interacted aminoacid	No. of interactions
1	MSH6	Acetophenone	2840	GLN-625, ASP-391	2
2	MSH6	Alpha-bisabolol	4414	LYS-853	1
3	MSH6	Alpha-terpineol	3278	LYS -853	1
4	MSH6	Alpha-thujene	-	-	-
5	MSH6	Borneol	2780	ARG -771	1
6	MSH6	Cis - cinnamaldehyde	3008	ASP-838	1
7	MSH6	Linalool	3634	LYS -832	1
8	MSH6	Styrene	-	-	-

Table 14. Docking of receptors with *Cyperus rotundus*

Sl. no.	Receptor Protein	Compound	Docking score	Interacted amino acid (in -kcal/mol)	No.of interactions
1	MLH1	2-propenoic acid	1428	THR -24	1
2	MLH1	carvone	2596	HIS -6, HIS -8	2
3	MLH1	Caryophyllene epoxide	-	-	-
4	MLH1	cyprotene	-	-	-
5	MLH1	Mytenol	2706	LEU -1	1
6	MLH1	Nookatone	3252	MET -4	1
7	MLH1	Sabinene	-	-	-
8	MLH1	verbenone	2834	LEU-1	1

Sl. no.	Receptor Protein	Compound	Docking score	Interacted amino acid	No. of interactions
1	MSH2	2-propenoic acid	2052	ASN -139	1
2	MSH2	carvone	3362	ASN-285	1
3	MSH2	Caryophyllene epoxide	3922	ARG -308	1
1	MSH2	cyprotene	-	-	-
5	MSH2	Mytenol	3144	SER-734	1
6	MSH2	Nookatone	3954	ILE -192	1
7	MSH2	Sabinene	-	-	-
8	MSH2	verbenone	3040	ARG -219	1

Sl. no.	Receptor Protein	Compound	Docking score	Interacted amino acid	No. of interactions
1	MSH6	2-propenoic acid	1970	GLN-1145, ASN-1110, GLY-1138	3
2	MSH6	carvone	3498	LYS -853	1
3	MSH6	Caryophyllene epoxide	-	-	-
4	MSH6	cyprotene	-	-	-
5	MSH6	Mytenol	3106	LYS -853	1

(*continued*)

Table 14. (*continued*)

Sl. no.	Receptor Protein	Compound	Docking score	Interacted amino acid	No. of interactions
6	MSH6	Nookatone	4368	SER -524	1
7	MSH6	Sabinene	-	-	-
8	MSH6	verbenone	3094	ASP -838	1

4372 kcal/mol with 1 interaction, Alpha-terpineol with a docking score of 3380 kcal/mol with 2 interactions, Borneol with a docking score of 2880 kcal/mol with 1 interaction, Cis-cinnamaldehyde with a docking score of 3190 kcal/mol with 2 interactions and Linalool with a docking score of 3740 kcal/mol with 2 interactions.

Also, it is seen that MSH6 docks with Acetophenone with a docking score of 2840 kcal/mol with 2 interactions, Alpha-bisabolol with a docking score of 4414 kcal/mol with 1 interaction, Alpha-terpineol with a docking score of 3278 kcal/mol with 1 interaction, Borneol with a docking score of 2780 kcal/mol with 1 interaction, Cis - cinnamaldehyde with a docking score of 3008 kcal/mol with 1 interaction and Linalool with a docking score of 3634 kcal/mol with 1 interaction. Again, it is seen that MLH1 docks with 2-propenoic acid with a docking score of 1428 kcal/mol with 1 interaction, carvone with a docking score of 2596 kcal/mol with 2 interactions, Mytenol with a docking score of 2706 kcal/mol with 1 interaction, Nookatone with a docking score of 3252 kcal/mol with 1 interaction and verbenone with a docking score of 2834 kcal/mol with 1 interaction. Also, it is seen that MSH2 docks with 2-propenoic acid with a docking score of 2052 kcal/mol with 1 interaction, carvone with a docking score of 3362 kcal/mol with 1 interaction, Caryophyllene epoxide with a docking score of 3922 kcal/mol with 1 interaction, Mytenol with a docking score of 3144 kcal/mol with 1 interaction, Nookatone with a docking score of 3954 kcal/mol with 1 interaction and verbenone with a docking score of 30402 kcal/mol with 1 interaction.

Further, it is seen that MSH6 docks with 2-propenoic acid with a docking score of 1970 kcal/mol with 3 interactions, carvone with a docking score of 3498 kcal/mol with 1 interaction, Mytenol with a docking score of 3106 kcal/mol with 1 interaction, Nookatone with a docking score of 4368 kcal/mol with 1 interaction and verbenone with a docking score of 3094 kcal/mol with 1 interaction. Again, MLH1 is seen to dock with 3-(2-Methoxyethyl)-1-nonanol with a docking score of 3906 kcal/mol with 1 interaction, bornylacetate with a docking score of 2730 kcal/mol with 1 interaction, Camphor with a docking score of 2342 kcal/mol with 1 interaction and Isobutyl phthalate with a docking score of 4218 kcal/mol with 1 interaction. Also, MSH2 docks with 3-(2-Methoxyethyl)-1-nonanol with a docking score of 4784 kcal/mol with 1 interaction, MSH2 with a docking score of bornylacetate 3436 kcal/mol with 1 interaction, Camphor with a docking score of 2914 kcal/mol with 2 interactions and Isobutyl phthalate with a docking score of 4580 kcal/mol with 1 interaction. Further, MSH6 docks with 3-(2-Methoxyethyl)-1-nonanol with a docking score of 5006 kcal/mol with 1 interaction, bornylacetate with a docking score of 3812 kcal/mol with 1 interaction, Camphor with a

Table 15. Docking of receptors with *Sonchus arvensis*

Sl. no.	Receptor Protein	Compound	Docking score (in –kcal/mol)	Interacted amino acid	No. of interactions
1	MLH1	3-(2-Methoxyethyl)-1-nonanol	3906	THR - 31	1
2	MLH1	3-methylnonane	-	-	-
3	MLH1	bornylacetate	2730	HIS-8	1
4	MLH1	Camphor	2342	THR - 24	1
5	MLH1	Isobutyl phthalate	4218	THR – 31	1
6	MLH1	p-Xylene	-	-	-
7	MLH1	Toluene	-	-	-
8	MLH1	Trans1,3-dimethylcyclopentane	-	-	-

Sl. No.	Receptor Protein	Compound	Docking score	Interacted amino acid	No. of interactions
1	MSH2	3-(2-Methoxyethyl)-1-nonanol	4784	SER - 825	1
2	MSH2	3-methylnonane	-	-	-
3	MSH2	bornylacetate	3436	GLN - 183	1
4	MSH2	Camphor	2914	ILE-192, LEU- 191	2
5	MSH2	Isobutyl phthalate	4580	THR - 677	1
6	MSH2	p-Xylene	-	-	-
7	MSH2	Toluene	-	-	-
8	MSH2	Trans1,3-dimethylcyclopentane	-	-	-

Sl. No.	Receptor Protein	Compound	Docking score	Interacted amino acid	No. of interactions
1	MSH6	3-(2-Methoxyethyl)-1-nonanol	5006	LEU -1263	1
2	MSH6	3-methylnonane	-	-	-
3	MSH6	Bornylacetate	3812	SER – 524	1
4	MSH6	Camphor	3028	LYS-853	1
5	MSH6	Isobutyl phthalate	5170	LYS -832, ASP-838	2
6	MSH6	p-Xylene	-	-	-
7	MSH6	Toluene	-	-	-
8	MSH6	Trans1,3-dimethylcyclopentane	-	-	-

docking score of 3028 kcal/mol with 1 interaction and Isobutyl phthalate with a docking score of 5170 kcal/mol with 2 interactions.

4 Conclusion

The taxonomy and functional information of *H. pylori* microbiome are identified. Again, as per docking studies and ADME analysis it is seen that phytocompounds Benzeneacetaldehyde, Betaionone, Dibutyl phthalate, Diacetonealcohol, Linalool, Eugenol, Alpha-terpineol, Cis-cinnamaldehyde, Acetophenone, Carvone, 2-propenoic acid, Camphor and Isobutyl phthalate 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 colon cancer.

References

1. Burkitt MD, Duckworth CA, Williams JM, Pritchard DM. Helicobacter pylori-induced gastric pathology: insights from in vivo and ex vivo models. Dis Model Mech. 2017;10(2):89-104. doi:https://doi.org/10.1242/dmm.027649
2. Kusters JG, van Vliet AH, Kuipers EJ. Pathogenesis of Helicobacter pylori infection. ClinMicrobiol Rev. 2006;19(3):449-490. doi:https://doi.org/10.1128/CMR.00054-05
3. Testerman TL, Morris J. Beyond the stomach: an updated view of Helicobacter pylori pathogenesis, diagnosis, and treatment. World J Gastroenterol. 2014;20(36):12781-12808. doi:https://doi.org/10.3748/wjg.v20.i36.12781
4. Zheng, D., Liwinski, T. &Elinav, E. Interaction between microbiota and immunity in health and disease. Cell Res 30, 492–506 (2020). https://doi.org/10.1038/s41422-020-0332-7
5. Kolligs, Frank T. "Diagnostics and Epidemiology of Colorectal Cancer." Visceral medicine vol. 32,3 (2016): 158-64. doi:https://doi.org/10.1159/000446488
6. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg. 2009;22(4):191-197. doi:https://doi.org/10.1055/s-0029-1242458
7. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointest Cancer Res. 2012;5(1):19-27.
8. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. Nat Rev Dis Primers. 2015;1:15065. Published 2015 Nov 5. doi:https://doi.org/10.1038/nrdp.2015.65
9. Boyle P, Langman JS. ABC of colorectal cancer: Epidemiology. BMJ. 2000;321(7264):805-808. doi:https://doi.org/10.1136/bmj.321.7264.805
10. Vasen, H., Tomlinson, I. & Castells, A. Clinical management of hereditary colorectal cancer syndromes.NatRevGastroenterolHepatol12,88–97(2015). https://doi.org/10.1038/nrgastro.2014.229
11. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology. 2010;138(6):2044-2058. doi:https://doi.org/10.1053/j.gastro.2010.01.054
12. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology. 2010;138(6):2073-2087.e3. doi:https://doi.org/10.1053/j.gastro.2009.12.064
13. 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.
14. 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.

15. 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.
16. Srinivasan, S., & Fredricks, D. N. (2008). The human vaginal bacterial biota and bacterial vaginosis. *Interdisciplinary perspectives on infectious diseases*, *2008*.
17. 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
18. 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
19. 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
20. 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.
21. https://www.molinspiration.com, Slovensky Grob, Slovakia
22. 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.
23. Hegde PL & Harini A, 2014, A text book of Dravyaguna Vijnana, Chaukhambha publications