IJBPAS, February, 2024, 13(2): 566-576

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

IN SILICO DRUG TARGET IDENTIFICATION AND IT'S CHARACTERISATION IN *ALISTIPES* SPP. INVOLVED IN HUMAN COLORECTAL CANCER: A PAN GENOME ANALYSIS APPROACH

GUPTA D* AND SAHA P

Department of Microbiology, The University of Burdwan, Golapbag More, Burdwan-713104, West-Bengal, India

***Corresponding Author: Dr. Debdoot Gupta: E Mail: dgupta@microbio.buruniv.ac.in**

Received 24th Dec. 2022; Revised 25th Jan. 2023; Accepted 8th May 2023; Available online 1st Feb. 2024

https://doi.org/10.31032/IJBPAS/2024/13.2.7726

ABSTRACT

Colorectal Cancer (CRC) is a global health problem and according to WHO (World Health Organization), causing as many as 690,000 deaths per year. Among the factors involved in the progression of CRC (diet, lifestyle, toxic habit, genetic and epigenetic effects) the gut microbiota play an important role in the health of the colon and any dysbiosis of it may have a significant role in the outcome of CRC. *Alistipes* is one of the bacterial genera involved in dysbiosis and formation of tumor. In this study, pan genome analysis of total nine species of the genus *Alistipes* was performed through the software BPGA. The core proteome set consists of 306 proteins of which 11 was found to be hypothetical. Out of those hypothetical proteins 5 proteins were found to be absent in human proteome and two of it also have no homologues in the beneficial gut bacteria *Lactobacillus spp*. So out of these two proteins, may be suitable for future drug target, one was found to be oligopeptide transporter protein. In silico sequence and structural analysis of this protein shows that it has 668 amino acids with a molecular weight of 70 kDa, theoretical pI of 8.40, instability index (II) of 25.12, aliphatic index is 115.99 and Grand average of hydropathicity (GRAVY) is 0.789. It's in silico predicted structure showed that it resembles the membrane bound protein with the combination of alpha-helices, beta-sheets and coils. Molecular Dynamics Simulation of this protein was also performed; this protein may be used as future drug target.

Keywords: Colorectal cancer, *Alistipes spp.*, Pan Genome, Drug target, Oligopeptide transporter

1. INTRODUCTION

Colorectal Cancer (CRC) is a global health problem according to WHO (World Health Organization) and was the leading cause of mortality positioned above stroke and coronary pathology. It positions third in incidence in men and second in women and one of the most frequently occurring forms of cancer worldwide causing as many as 690,000 deaths per year **[1]**. Different factors are involved for the progression of CRC and apart from the diet, lifestyle, toxic habit, genetic and epigenetic effects, the gut microbiota play an important role in the health of the colon and any dysbiosis of the gut microbiota may have a significant role in the outcome of CRC **[2]**. The disruption of the balance of good and bad bacteria of the human gut may play a role in the occurrence of inflammatory digestive diseases **[3]**, obesity **[4]**, colorectal adenoma and cancer **[5, 6]**. The association of gut microorganisms in causing and/or promoting CRC may be explained through a dynamic model known as 'Bacterial Driver-passenger' model **[7]**. The colonization of the germ free mice with the feces of healthy and CRC patients gives rise to different level of tumorigenesis and it was found that the susceptibility to tumorigenesis is linked with the initial component of the gut microbiota **[8]**. Some specific bacterial genera are found to be associated with the formation of tumor. Recent study has shown that some oral antibiotic promotes the growth of cancer causing bacteriome through dysbiosis of gut microbiota. Connection between oral antibiotic use, particularly beta-lactum antibiotic and increased risk of colon cancer also supports this fact **[9]**. It was also found the oral use of ampicillin/amoxicillin increased the risk of colon cancer whereas oral use of tetracycline reduces the risk of rectal cancer **[10]**. Antibiotic resistance of the cancer promoting bacteria is a major concern to be discussed; in particular, β-lactam antibiotic resistant fusobacteria is commonly found nowadays **[11]** though the normal gut microbiota is often sensitive to these antibiotics **[12]**. Therefore novel drug target identification and/or novel drug designing for those reported drug resistant bacteria involved in the progression of CRC may be a good starting point towards novel therapeutic strategy against colorectal cancer. There are various computational methods for novel drug target identification for drug resistant bacteria and pan genome analysis is one of them. Recently drug target identification from the core proteome set through pan genome analysis was performed for *Streptococcus gallolyticus* **[13]**, *Mycobacterium tuberculosis*

[14] and *Clostridium botulinum* **[15]**. The genus *Alistipes*, which is found as one component of the human gut microbiota from the colon cancer patient [16] was reported to be involved in inflammation, cancer and mental health **[17]**. In a recent experiment, it was revealed that the colonic tumor burden in germ free mice in which fecal microbiota of CRC patients was transplanted was reliant on some particular bacterial taxa one of which is the genus *Alistipes* **[18]**. In this study, the pan genome (consisting of core genome, dispensable genome and unique genome) analysis of *Alistipes spp* was performed and the core gene set of this bacterial species was explored for identification of drug target which may aid in the treatment of colorectal cancer in future. To the best of our knowledge this is the first report of drug target identification in the genus *Alistipes* through pan genome analysis.

2. MATERIALS AND METHODS

Collection of Genome sequences: The genome sequences of the different *Alistipes spp* are collected from the NCBI FTP server (ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/). The .gbff files of total nine species reported in NCBI at the time of this study were used as input for the pan genome analysis.

Pan genome analysis software: From the various pan genome analysis software in the web, the Bacterial Pan Genome Analysis Tool (BPGA) **[19]** was selected for this experimental work. This software has various functionality as of like core/pan/MLST (Multi Locus Sequence Typing) phylogeny, exclusive presence/absence of genes in specific strains, subset analysis, atypical G + C content analysis and KEGG & COG mapping of core, accessory and unique genes. This tool can be downloaded from the site sourceforge.net (http://sourceforge.net/projects/bpgatool/) through windows based installer.

In silico characterization of protein: For knowing the physical parameters of the protein, the PROTPARAM tool of ExPASy (https://web.expasy.org/protparam/) **[20]** is used. As the protein has no suitable homologues, Ab initio protein structure prediction approach was used through I-TASSER (https://zhanggroup.org/I-TASSER/) **[21]**.

3. RESULT AND DISCUSSION:

Collection of Genome Sequences: Total nine genome sequences of different *Alistipes* spp. were downloaded from the ftp sites for pan genome analysis which are shown in **Table 1**. The range of genome sizes and GC percentage are from 2.22 to 3.76 Mbp and 53.3 and 58.99 respectively. The highest protein count is seen in *Alistipes finegoldii*, species which was

reported to be present in the blood from colorectal cancer patient **[16]**.

Pan genome analysis of *Alistipes* Spp.: The components of pan genome of different *Alistipes* genomes are listed in **Table 2**.

From the data it was found that the pan genome contains 306 genes as core gene set which contains the genes responsible for the basic functionality of its activity. The highest number of accessory genes (2159), unique genes (1245) and exclusively absent genes (86) were found in:

Alistipes_finegoldii_DSM_17242, Alistipes_indistinctus_YIT_12060 and Alistipes_sp._58_9_plus respectively. From the COG distribution analysis of the core, accessory and unique genes, it was revealed that major portion of the core genes are involved in translation, ribosomal structure and biogenesis as these are the fundamental processes involved in basic biological activity. In KEGG distribution analysis, it was exposed that maximum number of core, unique and accessory genes are involved in carbohydrate metabolism. This finding is supported by the fact that many complex carbohydrates are degraded and fermented by the human gut microbiota in the large intestine to both yield basic energy salvage and impact gut health through produced metabolites **[22]**. Through mathematical interventions it was found that the pan genome and core genome of *Alistipes* genus follow power law and exponential law respectively and its pan genome is found to be open in spite of its restricted niche in the gut **(Table 3)**. This means when any new *Alistipes* genome will be included, it will add some new genes with the already stored collection of genes. It is noteworthy that the pan genome of *Helicobacter pylori*, the human gastric pathogen was also found to be open **[23]**.

Table 1: Features of different *Alistipes* genomes

Serial No	Name	Acc No.	Source	Size(Mb)	GC%	Gene Count	Protein Count
1.	*Alistipes finegoldii* DSM 17242	NC_018011.1	Appendix tissue	3.73	56.6	3259	3116
2.	*Alistipes indistinctus* YIT 12060	NZ_JH370371	Human Gut	2.86	54.8	2383	2303
3.	*Alistipes_inops*	NZ_JRGF01000001	Human feces	2.3	56.6	1775	1687
4.	*Alistipes putredinis*	NZ_ABFK02000000	Human gut	2.55	53.3	2303	2426
5.	*Alistipes shahii*	NC_021030	Human gut	3.76	57.2	3161	2919
6.	*Alistipes_sp._56_*11_human_gut_ metagenome_	GCA_001915255.1	Human gut	2.22	56.60	1945	1674
7.	*Alistipes sp. 58_9_*plus	MNQG01000000	Human gut	3.09	58.3	3321	2383
8.	*Alistipes sp.* CHKCI003	NZ_FCNT01000000	Human caecum	2.86	58.99	2537	2506
9.	*Alistipes_sp._*HGB5	GCA_000183485.2	gut	3.46	57.50	2987	2852

Table 2: Components of Pan Genomes of selected bacteria

Genome no.	Organism name	No. of core genes	No. of accessory genes	No. of unique genes	No. of exclusively absent genes
1	Alistipes_finegoldii_ DSM_17242	306	2159	403	0
2	Alistipes_indistinctus_ YIT_12060	306	731	1245	33
3	Alistipes_inops	306	1161	157	19
4	Alistipes_putredinis_ human_gut_metagenome_	306	1128	558	12
5	Alistipes_shahii_ WAL_8301	306	1336	795	58
6	Alistipes_sp._56_11_ human_gut_metagenome_	306	1107	200	47
7	Alistipes_sp._58_9_plus_ human_gut_metagenome	306	1385	401	86
8	Alistipes_sp._CHKCI003	306	1238	884	6
9	Alistipes_sp._HGB5	306	2134	351	1

Table 3: Fit laws for the pan and core genomes of *Alistipes spp*.

	PAN GENOME	CORE GENOME
Fit law	POWER	EXPONENTIAL
Equation	f(x)=a.x^b	f1(x)=c.e^(d.x)
Parameters	a= 2384.77	c= 2569.25
	b= 0.585887	d= -0.388594
Expected Size	8662	306
Estimated Size	8640.22	77.79

The parameter 'b' = 0.585887 , The pan genome is OPEN.

Therapeutic target identification from core gene set and it's in silico structural and functional characterization: From the core proteome set of 306 proteins, 11 hypothetical proteins were selected for functional analysis out of which 5 proteins were analyzed to be absent in human proteome and out of these 5 proteins 2 proteins also have no homologues in the beneficial gut bacteria (here *Lactobacillus acidophilus* was taken as a model of beneficial gut bacteria). So these 2 proteins may act as a good therapeutic target to be used in future against *Alistipes* spp. Of these two proteins, one protein was found to have the function of oligopeptide transporter (Acc No. EFR57896.1). Oligopeptide transporter is a group of proteins which are present in eukaryotes as well as in prokaryotes

and in bacteria this type of protein is not only involved in nutrient uptake but also in pathogenicity **[24]**. So this protein from *Alistipes spp*. was selected for its further in silico characterization.

Physicochemical properties of the selected core hypothetical protein: This protein has 668 amino acids with a molecular weight of 70 kDa and a theoretical pI of 8.40; high pH makes this protein active in alkaline range. Total numbers of positively and negatively charged amino acids are 38 and 41 respectively. The instability index (II) is computed to be 25.12 which classify the protein as a stable one and is supported by the fact that it has a half-life period of more than 10 hours in bacteria. The Aliphatic index is 115.99 which denotes its thermostability; Grand average of hydropathicity (GRAVY) is 0.789 which signifies its hydrophobic nature and that is supported by its membrane bound structure. All these physical parameters of his protein found from PROTPARAM are important for its better exploration through wet lab assay in future.

Structure related properties of the selected core hypothetical protein: The secondary structure of this protein shows that it has repeated helical structure with a high confidence score of 9 with intermittent coiled structure as got from the output of I-TASSER. The predicted solvent accessibility shows that the helical regions are buried and the first few amino acids are solvent accessible. The predicted normalized B-factor value is found to be less than 0 in all helical portions which indicates their stability in experimental structure. The topmost threading template used by I-TASSER for predicting the three dimensional structure of the query protein was the structure of eukaryotic voltage-gated sodium channel (PDB ID: 5XOM). The predicted structure of the hypothetical protein is given in the following figure **(Figure 1)**.

The structure also resembles with the typical membrane bound protein structure with the combination of alpha-helices, beta-sheets and coils. The stability of the structure was checked by Ramachandran plot analysis through RAMPAGE and it was found the 99% of the residues are in favorable regions. The protein was subjected to molecular dynamics simulation to verify its stability over nano second time level and it was found that the coiled regions (present in 200^{th}, 400^{th} and 600^{th} position) had more RMSD value than the remaining portions **(Figure 2)**. This may be due to their exposed nature in the structure.

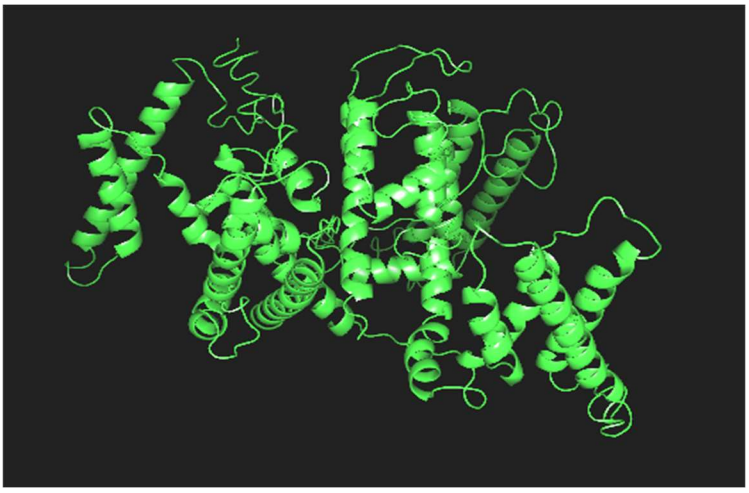

Figure 1: Predicted structure of the selected hypothetical protein of *Alistipes finegoldii*

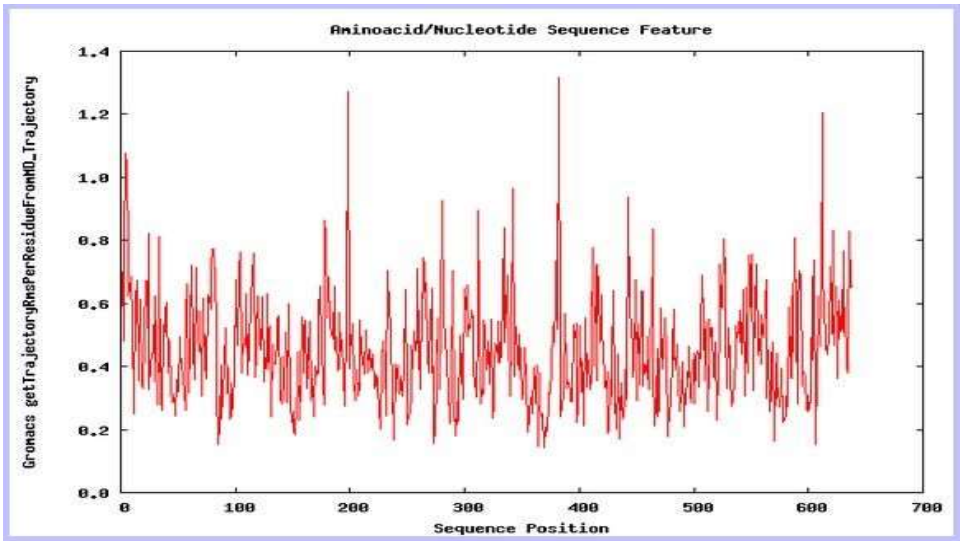

Figure 2: Residue wise RMSD value of the hypothetical protein of *Alistipes finegoldii*

4. CONCLUSION:

Pan genome contains the full repertoire of the genes present in any organism. In this study, pan genome analysis of the gut bacterium Alistipes *finegoldii* was done for identification of its drug target. For the pan genome analysis, total nine species of *Alistipes* genus were selected from the NCBI genome which showed variations in their genome size and GC percentage. As the highest protein count was seen in *Alistipes finegoldii* (bacterium reported to be found in the blood of colon cancer patient), this species was later analysed specially through different in-silico methods. The drug target was identified from the 306 core gene set, majority of which is involved in translation, ribosomal structure and biogenesis. In the metabolism of carbohydrate, amino acids and nucleotides, equal proportion of core, accessory and

unique genes are involved which signifies the importance of these functions in gut bacteria; on the other hand very few unique genes were found to be involved in cell motility. The core pan plot of this genus showed the 'open' and 'close' nature of pan genome and core genome respectively. For searching a universal and novel drug target for this genus, the core proteome was analyzed for the selection of hypothetical proteins which are conserved in this genus but absent in the host (*Homo sapiens*) as well as in beneficial gut bacteria (here *Lactobacillus acidophilus* was taken as a model of beneficial gut bacteria). Two hypothetical proteins passed these criteria and out of these two, one protein was analyzed to have the function of oligopeptide transporter and its physicochemical properties also put this protein as a good target for therapeutic interventions. Oligopeptide transporter protein was already found to be important as one of the possible therapeutic target in bacteria. So for exploring its functional and structural significance its three dimensional structure was predicted and its stability was verified by in silico methods. Molecular dynamics simulation study of the predicted structure showed that major portion of the structure is stable at nano second level except for the coiled regions. The predicted structure of this protein may be used for computer assistant drug designing against this bacterium in future.

5. ACKNOWLEDGEMENT:

I convey my special thanks to Dr. Sougata Roy, Assistant Professor, Ashoka University, Sonepat, Haryana for his valuable suggestion. I also convey my regards to the HOD, Department of Microbiology, The University of Burdwan for providing infrastructure for this computational work.

6. REFERENCES:

[1] Soerjomataram, J. Lortet-Tieulent, D.M. Parkin, J. Ferlay, C. Mathers, D.Forman, F. Bray, Global burden of cancer in 2008: a systematic analysis of disability-adjusted life-years in 12 world regions, Lancet 380 (2012) (2008), 1840–1850.

[2] Chen, Weiguang, Fanlong Liu, Zongxin Ling, Xiaojuan Tong, and Charlie Xiang. "Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer." *PloS one* 7, no. 6 (2012): e39743.

[3] Collins, Stephen M. "A role for the gut microbiota in IBS." *Nature reviews Gastroenterology & hepatology* 11.8 (2014): 497-505.

[4] Bajzer, Matej, and Randy J. Seeley. "Obesity and gut

flora." *Nature* 444.7122 (2006): 1009-1010.

[5] Dulal, Santosh, and Temitope O. Keku. "Gut microbiome and colorectal adenomas." *Cancer journal (Sudbury, Mass.)* 20.3 (2014): 225-231.

[6] Gao, Zhiguang, Bomin Guo, Renyuan Gao, Qingchao Zhu, and Huanlong Qin. "Microbiota disbiosis is associated with colorectal cancer." *Frontiers in microbiology* 6 (2015): 20, 1-9.

[7] Tjalsma, Harold, Annemarie Boleij, Julian R. Marchesi, and Bas E. Dutilh. "A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects." *Nature Reviews Microbiology* 10, no. 8 (2012): 575-582.

[8] Baxter, Nielson T., Joseph P. Zackular, Grace Y. Chen, and Patrick D. Schloss. "Structure of the gut microbiome following colonization with human feces determines colonic tumor burden." *Microbiome* 2, no. 1 (2014): 1-11. Wan, Q.-Y.; Zhao, R.; Wang, Y.; Wu, Y.; Wu, X.-T. Antibiotic use and risk of colorectal cancer: A meta-analysis of 412450 participants. Gut, 2020, 69, 2059–2060.

[9] Zhang, Jiajia, et al. "Oral antibiotic use and risk of colorectal cancer in the United Kingdom, 1989–2012: a matched case–control study." *Gut* 68.11 (2019): 1971-1978.

[10] Wan, Qian-Yi, Rui Zhao, Yong Wang, Yutao Wu, and Xiao-Ting Wu. "Antibiotic use and risk of colorectal cancer: a meta-analysis of 412 450 participants." *Gut* 69, no. 11 (2020): 2059-2060.

[11] Sorbara, Matthew T., and Eric G. Pamer. "Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them." *Mucosal immunology* 12.1 (2019): 1-9.

[12] Dudek-Wicher, R. K., A. Junka, and M. Bartoszewicz. "The influence of antibiotics and dietary components on gut microbiota", *Prz Gastroenterol*, 13(2018): 85–92.

[13] Qureshi, Nosheen Afzal, Syeda Marriam Bakhtiar, Muhammad Faheem, Mohibullah Shah, Ahmed Bari, Hafiz M. Mahmood, Muhammad Sohaib, Ramzi A. Mothana, Riaz Ullah, and Syed Babar Jamal. "Genome-based drug target identification in human pathogen *Streptococcus*

gallolyticus." *Frontiers in Genetics* 12 (2021): 564056.

[14] Dar, Hamza Arshad, Tahreem Zaheer, Nimat Ullah, Syeda Marriam Bakhtiar, Tianyu Zhang, Muhammad Yasir, Esam I. Azhar, and Amjad Ali. "Pangenome Analysis of *Mycobacterium tuberculosis* Reveals Core-Drug Targets and Screening of Promising Lead Compounds for Drug Discovery." *Antibiotics* 9, no. 11 (2020): 819.

[15] Bhardwaj, Tulika, and Pallavi Somvanshi. "Pan-genome analysis of *Clostridium botulinum* reveals unique targets for drug development." *Gene* 623 (2017): 48-62.

[16] Fenner, Lukas, Véronique Roux, Pascal Ananian, and Didier Raoult. "*Alistipes finegoldii* in blood cultures from colon cancer patients." *Emerging Infectious Diseases* 13, no. 8 (2007): 1260.

[17] Parker, Bianca J., Pamela A. Wearsch, Alida CM Veloo, and Alex Rodriguez-Palacios. "The genus Alistipes: gut bacteria with emerging implications to inflammation, cancer, and mental health." *Frontiers in immunology* 11 (2020): 906.

[18] Baxter, Nielson T., Joseph P. Zackular, Grace Y. Chen, and Patrick D. Schloss. "Structure of the gut microbiome following colonization with human feces determines colonic tumor burden." *Microbiome* 2, no. 1 (2014): 1-11.

[19] Chaudhari, Narendrakumar M., Vinod Kumar Gupta, and Chitra Dutta. "BPGA-an ultra-fast pan-genome analysis pipeline." *Scientific reports* 6.1 (2016): 1-10.

[20] Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. "Protein identification and analysis tools on the ExPASy server." *The proteomics protocols handbook* (2005): 571-607.

[21] Yang, Jianyi, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, and Yang Zhang. "The I-TASSER Suite: protein structure and function prediction." *Nature methods,* 12(1) (2015): 7-8.

[22] Chassard, Christophe, and Christophe Lacroix. "Carbohydrates and the human gut microbiota." *Current Opinion in Clinical Nutrition & Metabolic Care* 16.4 (2013): 453-460.

[23] Ali, Amjad, Anam Naz, Siomar C. Soares, Marriam Bakhtiar, Sandeep Tiwari, Syed S. Hassan, Fazal Hanan et al. "Pan-genome analysis of human gastric pathogen H. pylori: comparative genomics and pathogenomics approaches to identify regions associated with pathogenicity and prediction of potential core therapeutic targets." *BioMed research international* 2015 (2015).1-17.

[24] Garai, Preeti, Kasturi Chandra, and Dipshikha Chakravortty. "Bacterial peptide transporters: Messengers of nutrition to virulence." *Virulence* 8.3 (2017): 297-309.