Pak. J. Biochem. Mol. Biol., 2022, 55(1):12-30

Research paper

Pan-genomic analysis of *Campylobacter jejuni* confers hypervariable sequences and functional genes involved in pathogenicity

S. M. Iqbal Azimuddin[1,2,3,4*], Talyha Khalid [1], Sayyada Ghufrana Nadeem[1]

[1]Department of Microbiology, Jinnah University for Women, Karachi-74600, Pakistan

[2]Xinjiang Key Laboratory of Environmental Pollution and Bioremediation, Xinjiang Institute of Ecology and Geography, Urumqi, Xinjiang, China

[3]University of Chinese Academy of Sciences, Beijing 100864, China

[4]Department of Biosciences, Mohammad Ali Jinnah University, Karachi-75400, Pakistan

*corresponding author: Dr. S. M. Iqbal Azimuddin (iqbal.azimuddin@jinnah.edu)

ABSTRACT

The bacterial Pan-genomic analysis gives a comprehensive view of genes present in strains. It offers detailed views of genetic diversity and adaptation among closely related strains such as duplication, loss of genes, and horizontal genes transfer mechanism. *Campylobacter jejuni* is associated with various food-borne illnesses worldwide. In this study, the Pan-genomic Analysis Pipeline tool (PGAP) was used to perform genomic analysis of *Campylobacter jejuni* to find functional gene distribution that might be responsible for normal gene functioning and virulence. A total of 56,455 whole proteins (Wp_Count) were present in the strains of *Campylobacter jejuni* and the subdivision of these Wp_Count in the core, shared and unique genome were 44,845, 10,981, and 629. In the core genome, 211 genes were found to be involved in pathogenicity. Flagellar assembly protein (FliW) and two-compartment system (TCS) flagellar assembly protein response regulator was found in all 35 strains which play an important role in host colonization. The other important functional genes were also present, which are reportedly involved in host immune invasion and pathogenicity. In the shared genome, 11 genes were found against selected query proteins and cluster analysis revealed hypervariable sequences were present in gene coding flagellum. Phylogenetic analyses revealed the evolution in *Campylobacter jejuni* flaA and flaB genes occurred at the same time. Phylogenetic analysis also discloses that the evolution of the *Campylobacter jejuni* strains was acquired by the high rate of recombination possibly due to horizontal gene transfer (HGT) showed variability and diversity in allelic genes. There was no unique gene found with reference to the query protein. This study provides baseline data to find a possible solution for improved treatment and overcome the resistance pattern within the specie. Apart from limited time and resources in the future, our focus is to advance the current study to do protein modeling and targeted drug development.

KEYWORDS *Campylobacter jejuni*, Horizontal gene transfer, Pan-genomic analysis, Phylogenetic analysis, Protein modelling

INTRODUCTION

Campylobacter species including *Campylobacter jejuni* are associated with a range of gastrointestinal conditions, including inflammatory bowel disease (IBD), colorectal cancer, and Barrett's esophagus. The role of disease development of these clinical conditions is unidentified. Genome annotations of several *Campylobacter jejuni* strains are not made of pilus-like open reading frames [1,2]. The fibronectin-binding domain of *CadF* consists of amino acids (*FrLS*) protein which represents a novel fibronectin-binding motif which required by *Campylobacter jejuni* for *in vitro* invasion [3]. ATP-binding cassette (ABC) transporters factor plays an important role to adhere with in the host cells and contribute in the pathogenesis. *Campylobacter*iosis is known to be self-limiting, and it is possible to treat the lost fluids and electrolytes [4,5]. However, if not treated timely the situation got reversed or may cause prolonged enteritis, septicemia, or severe intestinal complications in immune compromised patients, in this case antibiotics can be useful. According to the World Health Organization WHO an increasing problem related to public health is antibiotic resistance among *Campylobacters spp* specially in *Campylobacter jejuni* and *Campylobacter coli.* The antibiotic resistance in *Campylobacters* is increasing against macrolides erythromycin, fluoroquinolones, and tetracycline, which are considered as the most persistent or recurring prescribed antibiotics against *Campylobacter*. However, an alternative treatment such as gentamicin is also useful [6].

Genomes which represent the same species can be extremely different and vary in sizes already discovered by pulse field gel electrophoresis in 1990s [7]. Whole genomic sequence comparison of isolates showed a large degree of intra-species variability [8,9] . A single bacterial isolate does not contain the ultimate genetic selection of its phylogenetic heredity. However, a part of pan-genes is unique to this isolate. Generally, the Pan-genomic of a bacteria's subset creates a union of three unique sets of genes such as shared genes, core genes, and unique genes. These three sets of genes reveal different properties for example the distribution of classes and the number of genes included [10,11,12,13,14]. Hence, accessory, core and unique genes are beneficial for various applications to offer important information regarding the group of investigated genomes. Therefore, the main objective of this study was to do pan-genomic analysis to discover and examine frequent genomic features that are intrinsic to composed species in core, dispensable and unique genome by exploring its genetic factors involved in pathogenicity to construct phylogenetic relationships.

MATERIALS AND METHODS

Selection of *Campylobacter jejuni* query proteins

A total of "50" proteins were selected involved in pathogenicity based on literature survey. Fasta sequences of the mentioned proteins were downloaded from Universal Protein Resource i.e. UniProt db. This database is openly available and provides comprehensive information about proteins.

Preparation of input data for PGAP Analysis

The genome sequences of *Campylobacter jejuni* strains was download with the help of wget command. Respected "latest assembly version" folders contained each strains data files including, Feature_table.txt.gz, Genomic.fna.gz and Protein.faa.gz. These files were copied to "campy-out" folder with the help of python script "copy2onedirectory.py". These files were decompressed and copied via Gzip 1.3.12 tool (http://ftp.gnu.org/gnu/gzip/gzip-1.3.12.tar.gz). These files were used to convert into special format by PGAP (Feature_table.txt) followed by annotation files (.function), Genomic.fna into Nucleotide sequences (.nuc) and Protein.faa into Protein sequences (.pep). Conversion of above file was done with the help of perl converter and syntax was generated followed by Perl data conversion and protein blast for multiple strains of *Campylobacter jejuni* against query protein. Blast (Basic Alignment Search Tool) and protein blast is the variant of blast represented as "Blastp". Blastp tool was used to compare

query protein sequences and subject sequences [15]. Blastp programmed were downloaded and installed on windows "ncbi-blast (2.7.1) ncbi-blast+2.7.1+-x64-win64.tar.gz".

PGAP Syntax Generation

PGAP syntax was generated by using python script run by defining the path of input and output directory. In this study GF method were used with default parameters. Once this command was executed error file was generated which showed that the length of wp_number "protein sequences" and corresponding nucleotide sequence were not consistent so before proceeding further error were needed to be corrected.

PGAP Output Files

Pangenomic analysis pipeline tool creates five different types of files. These files were further used for genomic analysis of *Campylobacter jejuni* strains. These files also contain sub file the major five file which are generated after PGAP run as, Orthologous cluster file, Pan-genomic data file, Coding sequence (CDS) variation for evolution file, Pan based UPGMA file and Orthologous Cluster orthologous gene (COG) Distribution file for: core, specific and dispensable. Among these output files we used Orthologous cluster file, Pan-genomic data file and Pan based UPGMA file for further analysis. Distribution of core genes, Distribution of shared genes and Distribution of unique genes was done by using Orthologous cluster file.

Cluster Analysis Reveals Hypervariable Sequences

Orthologous gene cluster.txt file was previously used to separate orthologous gene cluster of "35" strains of *Campylobacter jejuni* into core, strain specific, and dispensable genes. Finding cluster sequences related to evolution, those which have same taxonomic unit, segment related to conserved and non-conserved sequences. Multiple sequence alignment was performed by selecting genes that have hypervariable sequences. The gene ID wp_014516840 was found in the shared genome of *Campylobacter jejuni* in (cluster ID-19) and all homologous protein sequence was exported for phylogenetic analysis. Flagellin cluster found to have most of the hypervariable sequence than any other homologous clusters. Pan genomic analyses was done by using PanGP software and Dendroscope (version 3.6.3) software was used to construct phylogenetic tree.

Selection of Pooled Genes

Extraction of pooled genes was done by using three files i.e., core.txt, shared.txt and unique.txt segregated earlier from orthologous cluster file. Genes were searched with reference to selected strains by using python script "wp_enquire_defsyntax". This script returns information about common genes in each strain by giving a list of IDs found in strains and not found in strains (GCF numbers) along with their functions and sequence length.

RESULTS

Strain selection based on sequence homology

Selection of query protein sequences and strains protein sequences was done on the basis of protein blast (Blastp) hits on 90-100% sequences homology. Among "50" query proteins sequences and "114" genome sequences with 185,398 genes. Total "35" Campylobacter jejuni strains and "33" query proteins were selected and was further subjected to study.

Pangenomic Analysis of Orthologous cluster and Common Genes

The strains vary in their genome sizes from (1.6 to 1.8 Mbp) and their number of predicted protein-coding sequences. Orthologous cluster file was used to segregate into core, shared and unique genome of *Campylobacter jejuni*. Analysis of 35 strains of *Campylobacter jejuni* reveals the cluster and genes count were calculated (Figure 1).

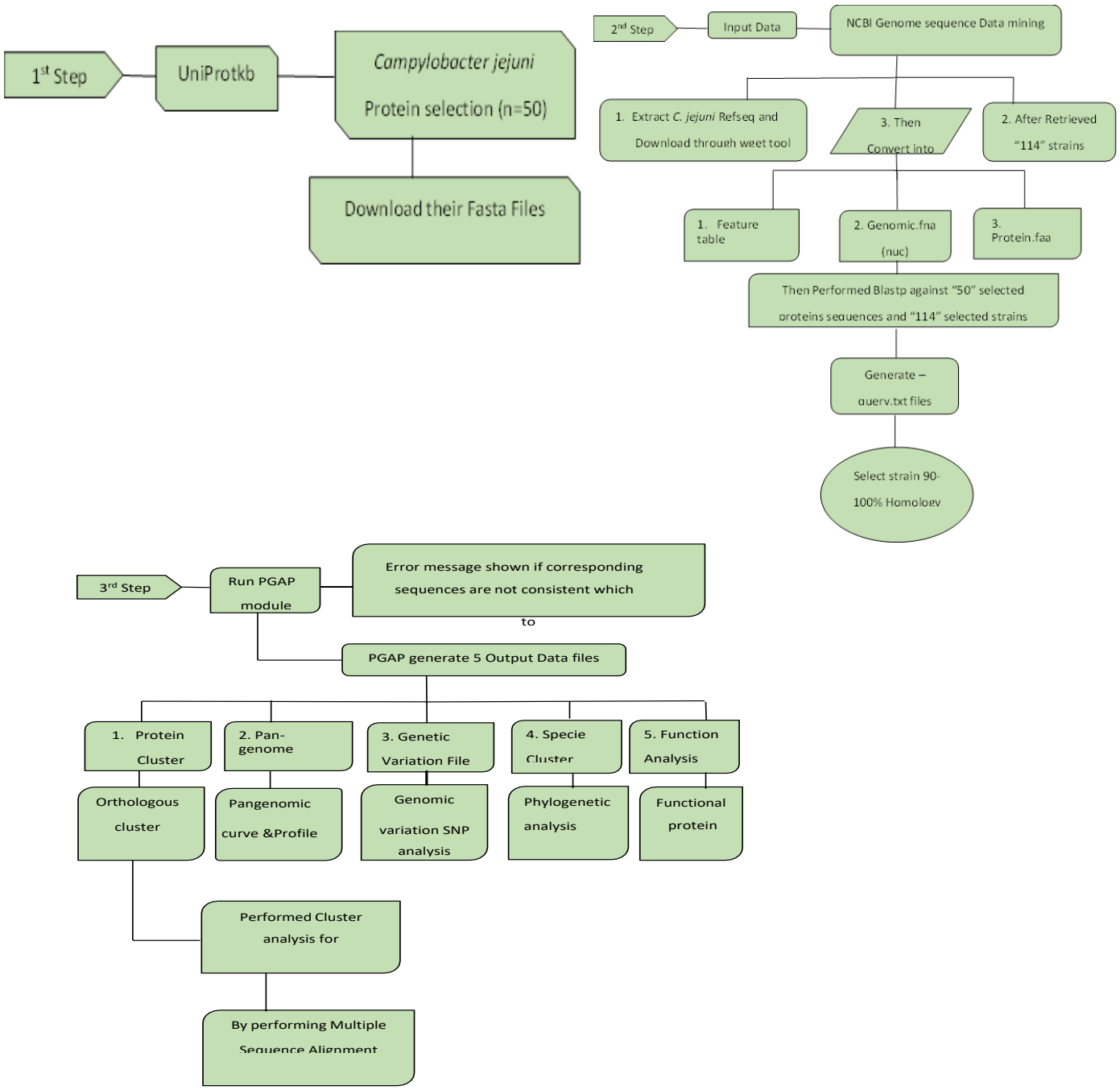

Figure 1: Flow Diagram of Pan-genomic Analysis of *Campylobacter jejuni* strain.

Total 56,455 genes and 2,765 clusters were found in 35 strains of *Campylobacter jejuni.* Whole proteins genes 56,455 were divided into three group's core, shared ad unique. Core genome encompassed 44,845 genes while shared genome consists of 10,981 genes and genes that are species specific includes 629 genes. In cluster orthologous analysis we found 2,765 clusters which further divided into core shared and unique genome. In core genome 1276 cluster was present; only 876 genes were shared by all strains under this study while 622 genes were found in species specific strains shown in (Figure 2). Although, the core genome cluster consists of 45.80%, only 31.60% of the clusters were shared by some strains but not all and 22.80% of the cluster were present in unique genome, these percentages showed high level of genomic diversity in the strains of *Campylobacter jejuni* genome. Among 222 genes 25 genes are present in the majority of the strains which are associated with the pathogenicity were listed in (Table 1). Although according to results no unique gene was found with reference to query protein indicating there was no unique query protein present in single strain of *Campylobacter jejuni*.

Cluster-based Analysis reveals Hypervariable sequences in shared genome.

Under this study Pangenomic comparison of thirty-five strains of *Campylobacter jejuni* cluster also reveals single hypervariable sequences in shared strains, Cluster ID-19 showed most of the hypervariable sequences in Flagellin genes due to homopolymeric tracts. These variable sequences also contain single nucleotide polymorphism, repeated sequence in the genome of *Campylobacter jejuni* provide rapid variation and also give great advantage to colonize in host immune system. In the genome of *Campylobacter jejuni* synonymous mutation was about "59421" of total genome i.e. (41.40%), non-synonymous mutation was "61489" of the total genome of (42.8%) which is lesser than the synonymous mutation and only "22608" in total number of indel mutation were found in whole genome of *Campylobacter jejuni* i.e. (15.7%) enlist in (Table 2). Most of the hypervariable regions were found in gene coding flagellum is shared genome in multiple strains of the *Campylobacter jejuni* sequences were shared. Phylogenetic analysis of hypervariable sequences showed distances during evolution (Figure 3). The phylogenetic analyses revealed among 35 strains total of 25 strains contained these genes and 10 strains lack these genes (Table 3).

Table 1: List of 25 pooled genes present in multiples strains of *Campylobacter jejuni* along with Pan-genomic genes distribution

No.	Gene number	Sequence length	Pan-genomic distribution	GCF found in 35 strains (%)	Genes Function
1	WP_002852982	129	Core	35 (100%)	Flagellar assembly factor fliw
2	WP_002866134	130	Core	35 (100%)	Two-component system response regulator
3	WP_002852579	90	Core	29 (82.85%)	30S ribosomal protein S15
4	WP_002853252	259	Core	28 (80%)	Major cell-binding factor
5	WP_002864738	232	Core	28 (80%)	Flagellar L-ring protein
6	WP_002865901	262	Core	26 (74.28%)	Chemotaxis protein methyltransferase
7	WP_002865902	242	Core	25 (71.42%)	Polar amino acid ABC transporter ATP-binding protein
8	WP_002866273	376	Core	23 (65.71%)	UDP-4-amino-4,6-dideoxy-N-acetyl-beta-L-altrosamine transaminase
9	WP_002866281	274	Core	23 (65.71%)	UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase
10	WP_002866200	165	Core	22 (62.85%)	Signal transduction histidine kinase
11	WP_002932689	200	Core	22 (62.85%)	Sugar transferase
12	WP_002932692	386	Core	22 (62.85%)	Aminotransferase DegT
13	WP_002932693	590	Core	22 (62.85%)	Polysaccharide biosynthesis protein
14	WP_002933046	127	Core	22 (62.85%)	Sec-independent protein translocase tatc

15	WP_002933204	461	Core	22 (62.85%)	Bifunctional-heptose-7-phosphatekinase/heptose-1 phosphate adenyltransferase
16	WP_002933464	128	Core	22 (62.85%)	Flagella export chaperone flis
17	WP_002933602	334	Core	22 (62.85%)	UDP-N-acetylglucosamine 4,6-dehydratase (inverting)
18	WP_002934195	543	Core	22 (62.85%)	CTP synthase
19	WP_002934339	372	Core	22 (62.85%)	Lipoprotein
20	WP_032598731	196	Core	22 (62.85%)	Acetyltransferase
21	WP_072225867	359	Core	22 (62.85%)	Glycosyltransferase family 4 protein
22	WP_002932672	309	Core	22 (62.85%)	Glycosyltransferase family 2 protein
23	WP_002932675	365	Core	22 (62.85%)	Glycosyltransferase
24	WP_002866201	459	Shared	20 (57.14%)	Biotin synthase
25	WP_002932677	713	Core	20 (57.14%)	Peptide-binding protein

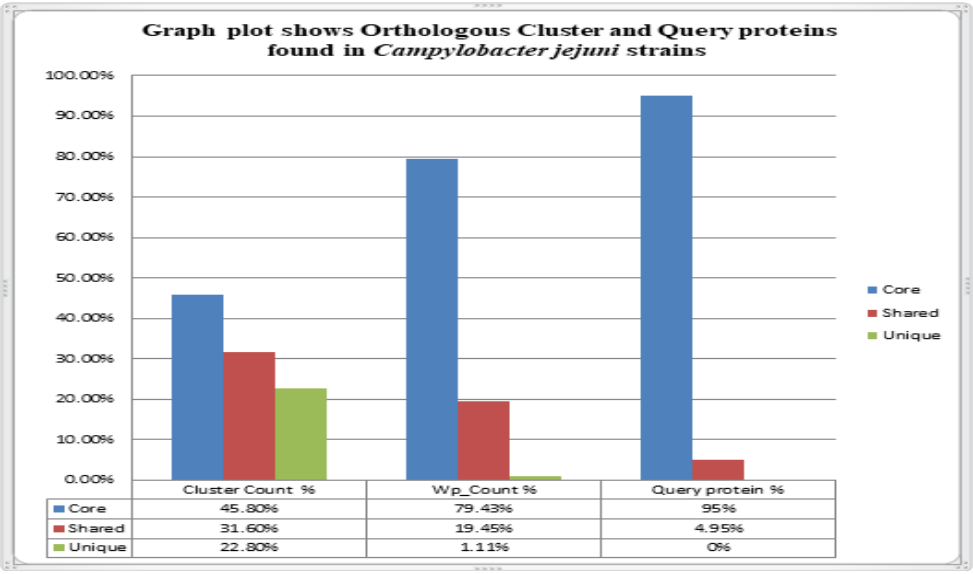

Figure 2: Graph plot between cluster count, Wp Count and query protein count showing that maximum gene and cluster were found in core genome.

Table 2: Variation found in Thirty-five strains of *Campylobacter jejuni* due to high rate of mutation

S. No	Variation Type	Found in Number	Percentage's
1	Synonymous mutation	59421	41.40%
2	Non-synonymous mutation	61489	42.84%
3	Indel mutation	22608	15.75%
Total		143518	99.99%

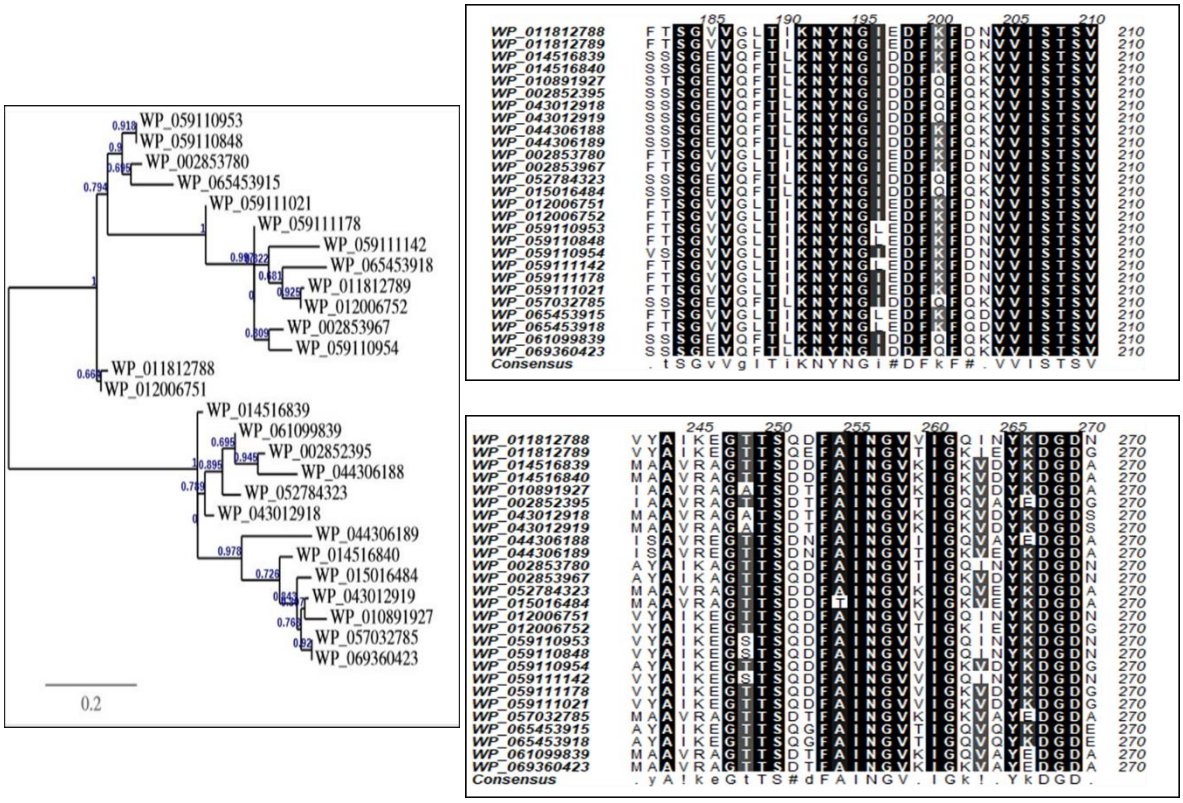

Figure 3: Phylogenetic analysis of *Campylobacter jejuni* hypervariable sequence showed flagellin genes are closely related among different strains

Table 3: *Campylobacter jejuni* 35 strains contain flagellin genes shared by 25 strains.

S. No	Genome Assembly Number	Pan Distribution	Flagellin Genes
1	GCF_000015525	shared	WP_011812788,WP_011812789
2	GCF_000025425	shared	WP_014516839,WP_014516840
3	GCF_000830805	shared	WP_002852395,WP_010891927
4	GCF_000830845	shared	WP_002852395,WP_010891927
5	GCF_000835285	shared	WP_002852395,WP_010891927
6	GCF_000835345	shared	WP_043012918,WP_043012919
7	GCF_000934305	shared	WP_044306188,WP_044306189
8	GCF_001299565	shared	WP_002853780,WP_002853967
9	GCF_001314285	shared	WP_052784323,WP_015016484
10	GCF_001412295	shared	WP_012006751,WP_012006752
11	GCF_001506185	shared	-
12	GCF_001506225	shared	-
13	GCF_001506265	shared	WP_059110953,WP_059110954
14	GCF_001506305	shared	WP_059110848
15	GCF_001506405	shared	WP_059110953,WP_059110954
16	GCF_001506485	shared	-

17	GCF_001506525	shared	-
18	GCF_001506565	shared	-
19	GCF_001506605	shared	WP_059110953,WP_059110954
20	GCF_001506685	shared	-
21	GCF_001506725	shared	-
22	GCF_001506765	shared	WP_059110953,WP_059110954
23	GCF_001506805	shared	-
24	GCF_001506845	shared	WP_059110953,WP_059111142
25	GCF_001506925	shared	WP_059110953,WP_059110954
26	GCF_001506965	shared	-
27	GCF_001507005	shared	WP_059111178
28	GCF_001507045	shared	WP_059111178
29	GCF_001507085	shared	WP_059111021
30	GCF_001507125	shared	-
31	GCF_001507165	shared	WP_059110953,WP_059111142
32	GCF_001507205	shared	WP_059110953,WP_059110954
33	GCF_001563565	shared	WP_061099839,WP_057032785
34	GCF_001686905	shared	WP_065453915,WP_065453918
35	GCF_001721965	shared	WP_061099839,WP_069360423

Key: (-) not found in strain

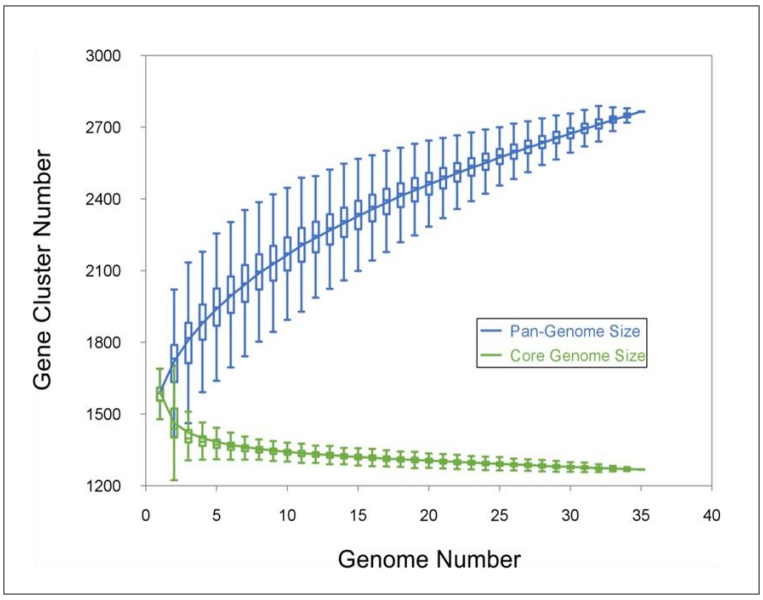

Figure 4: Pan-genome Profile curve indicating the characteristics of Pan-Genome size increase and Core Genome size decrease by addition of new genes.

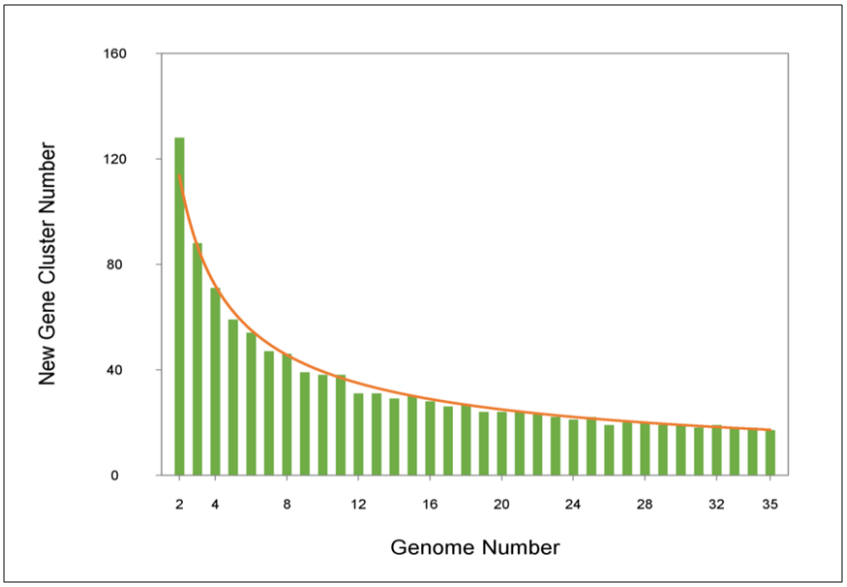

Figure 5: New gene curve of thirty-five strains of *Campylobacter jejuni* showed open genome.

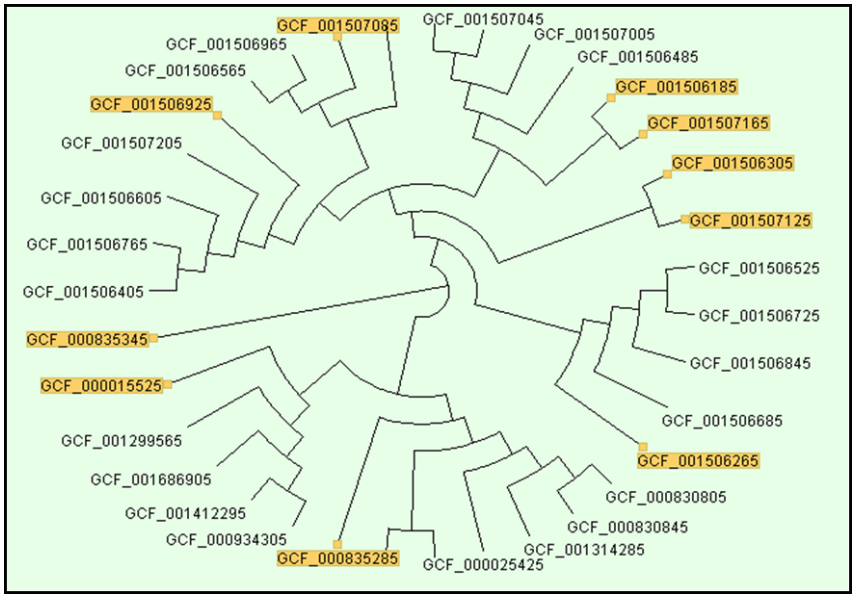

Figure 6. Phylogenetic analysis showed that the strains are closely related and the highlighted strains of *Campylobacter jejuni* were possibly the most common ancestor of other strains.

Table 4: List of selected strains of *Campylobacter jejuni* their common name and genome size.

S. No	Assembly	Common Name	Genome size	Total Genes
1	GCF_000835345	*Campylobacter jejuni* subsp. *Jejuni* strain 01-1512	1742364 bp	1862 genes
2	GCF_000835285	*Campylobacter jejuni* subsp. *Jejuni* strain 00-1597	1742047 bp	1694 genes
3	GCF_000025425	*Campylobacter jejuni* subsp. *jejuni* IA3902	1635 045 bp	1670 genes
4	GCF_000015525	*Campylobacter jejuni* subsp. *jejuni* 81-176	1616554 bp	1659 genes
5	GCF_001563565	*Campylobacter jejuni* strain RM3194	1651183 bp	1653 genes
6	GCF_001686905	*Campylobacter jejuni* subsp. *jejuni* strain RM1285	1675112 bp	1638 genes
7	GCF_001506685	*Campylobacter jejuni* strain CJ677CC538	1675952 bp	1636 genes
8	GCF_000830845	*Campylobacter jejuni* subsp. *jejuni* NCTC 11168-mfK12E5	1641469 bp	1635 genes
9	GCF_001506845	*Campylobacter jejuni* strain CJ677CC523	1667224 bp	1634 genes
10	GCF_000830805	*Campylobacter jejuni* subsp. *jejuni* NCTC 11168-Kf1	1641480 bp	1633 genes
11	GCF_001721965	*Campylobacter jejuni* subsp. *Jejuni* strain MTVDSCj13 chromosome	1684409 bp	1632 genes
12	GCF_001506725	*Campylobacter jejuni* strain CJ677CC014	1671231 bp	1631 genes
13	GCF_001506265	*Campylobacter jejuni* strain CJ677CC073	1672553 bp	1631 genes
14	GCF_001506525	*Campylobacter jejuni* strain CJ677CC531	1666051 bp	1630 genes
15	GCF_001506305	*Campylobacter jejuni* strain CJ677CC526	1667602 bp	1619 genes
16	GCF_001507125	*Campylobacter jejuni* strain CJ677CC024	1661567 bp	1614 genes
17	GCF_001299565	*Campylobacter jejuni* subsp. *jejuni* strain RM3197	1664565 bp	1609 genes
18	GCF_001506565	*Campylobacter jejuni* strain CJ677CC059	1632509 bp	1586 genes
19	GCF_001506225	*Campylobacter jejuni strain CJ677CC534*	1633831 bp	1580 genes
20	GCF_001507165	*Campylobacter jejuni strain CJ677CC525*	1642039 bp	1579 genes
21	GCF_001506185	*Campylobacter jejuni strain CJ677CC519*	1642794 bp	1579 genes

22	GCF_001506605	*Campylobacter jejuni strain CJ677CC033*	1634109 bp	1579 genes
23	GCF_001506405	*Campylobacter jejuni strain CJ677CC041*	1636128 bp	1578 genes
24	GCF_001506485	*Campylobacter jejuni strain CJ677CC532*	1643038 bp	1578 genes
25	GCF_001507005	*Campylobacter jejuni strain CJ677CC013*	1642694 bp	1578 genes
26	GCF_001506965	*Campylobacter jejuni strain CJ677CC047*	1635237 bp	1577 genes
27	GCF_001506805	*Campylobacter jejuni strain CJ677CC527*	1643938 bp	1576 genes
28	GCF_001507045	*Campylobacter jejuni strain CJ677CC522*	1642497 bp	1575 genes
29	GCF_001507085	*Campylobacter jejuni strain CJ677CC008*	1624545 bp	1575 genes
30	GCF_001507205	*Campylobacter jejuni strain CJ677CC034*	1637207 bp	1575 genes
31	GCF_001506765	*Campylobacter jejuni strain CJ677CC085*	1636377 bp	1573 genes
32	GCF_001506925	*Campylobacter jejuni strain CJ677CC539*	1636223 bp	1569 genes
33	GCF_001412295	*Campylobacter jejuni strain CJM1cam*	1616662 bp	1564 genes
34	GCF_000934305	*Campylobacter jejuni subsp. Jejuni strain 35925*	1612378 bp	1546 genes
35	GCF_001314285	*Campylobacter jejuni strain RM1285*	1635803 bp	1467 genes

Pan-Genomic Profile Curve

The presence of shared and unique genes made size of Pan-genome relatively higher between strains of *Campylobacter jejuni* indicating addition of new genes (Figure 4 and 5). The size of Pan-genome increases by the addition of new genes and core genome size decrease with addition of new strains, addition of new genes expressed as the *Campylobacter jejuni* have an open genome. Relationship between number of cluster (pan-genome) and core genome were calculated by PanGP software shown in (Figure 5 and Figure 6). The relation between Pan-genome size (y) and the number of genome (x) was $y= 408.0\ x^{0.3779} + 1187.0601$ (R^2 = 0.999981). The genome of *Campylobacter jejuni* is open because as the Pan-genome size increases as the number of genomes sequenced is also increased. PGAP also calculated the relation between core genome size (y) and genome number (x) as $y= 301.635\ x^{(-0.205*x)} + 1289.832$ (R^2 = 0.92600306).

Fitting for Pan-genome Profile

Function model: $Y = Ax^B + C$, Y = Pan-genome size, X = genome number, A, B, C: fitting parameters

1. y = 408.000995316928 *x**0.377999999999999 + 1187.06016883144
2. R-square = 0.999981774124527
3. A = 95% confidence interval: (408.000995316928 - 0.0982639070533244, 408.000995316928 + 0.0982639070533244)
4. C = 95% confidence interval: (1187.06016883144 - 0.361847611703211, 1187.06016883144 + 0.361847611703211)

Fitting for Core Genome Profile

Function model: Y = AxBx +C, Y = Pan-genome size, X = genome number, A, B, C: fitting parameters

1. y = 301.635861542283 *exp (-0.205 * x) + 1289.83280236024 R-square = 0.926003065427518
2. R-square = 0.926003065427518
3. A = 95% confidence interval: (301.635861542283 - 12.4995410706444, 301.635861542283 + 12.4995410706444)
4. C = 95% confidence interval: (1289.83280236024 - 1.11551150694153, 1289.83280236024 + 1.11551150694153)

Fitting Curve for New Gene Profile

Model: $Y = Ax^B$, Y: new gene size, X: genome number, A, B: fitting parameters

1. A: 179.808 B: -0.66
2. A = 95% Confidence Interval: 177.03 < A<182.62
3. B = 95% Confidence Interval: -0.66 + 0,

4. R Squared: 0.992465

Pan-genome based Phylogenetic Analysis of *Campylobacter jejuni (Figure 6)*

PGAP creates a phylogenetic tree based on two types of data. The first includes a gene distance matrix and nucleotide sequence of core gene clusters. The second includes the mutation and indel variations in the core gene cluster. Two algorithms, namely neighbor-joining (NJ) and UPGMA, are present in our results file. Phylogenetic relationship between Thirtyfive strains of *Campylobacter jejuni* was also constructed in this study by analysing Pan-genome based UPGMA. *Campylobacter jejuni* genomic evolution and their distribution were identified with the support of phylogenetic tree about its branch nodes (Figure 7) the strain GCF_000835345 (*Campylobacter jejuni* subsp. Jejuni strain 01-1512) is the highest genome size among all strains i.e. 1742,364 bp with 1862 genes extracted in total listed in (Table 4) was consider to be divergent strain among other strains. Almost 90 to 100% protein which is selected for this study is found under this

particular strain and having 168 extra genes from other strains. The distribution of other strains is under the branch of this strain GCF_000835345 and the branches of phylogenetics also reveal that the other strains are closely related to one and other.

DISCUSSION

Shared genome cluster analysis of *Campylobacter jejuni* strains

Shared genome cluster analysis of *Campylobacter jejuni* strains sequences was performed which directed toward the several hypervariable region or identification of plastic regions in flagellin genes [16]. The process of variability is mediated by high mutation rate and reversible mutation in the single sequence repeat referred as phase variation (PV). The wide most striking variability was found in shared genome during clusters analysis in flagellin genes. In cluster, encoding proteins involved in the biosynthesis of surface structure such as LOS, Flagellin, in post translational modification through glycosylation and as well as in the capsular gene which encodes proteins. The divergence in genes not only represents the variability among the predicted gene but also predict the presence or absence of specific gene products. Similar divergence was also detected in strain 81-176 in region intricate modification and synthesis of these surface structure and previous observation also confirmed of limited sequences analysis of some of these loci [17].

Pan-genome Analyses

Pan-genome size increases with the addition of genomes (strains) but the core genome size decreased. Hence this curve inferred that the pan-genomic curve of *Campylobacter jejuni* strains was open. While new gene curves indicated the frequency of new gene clusters within 35strains of *Campylobacter jejuni*. The genome curve (red) indicates the number of new genes with the increase of *Campylobacter jejuni* genome. Similar result was found in literature [18]. Other study also revealed the species of *Campylobacter* has open pan-genome. Many functional genes were involved in processes such as metabolism, defence and virulence mechanism. These genes likely to transfer through horizontal gene transfer play an important role in specie genetic diversity [19].

Genetic Variation in the Genome of *Campylobacter jejuni*

Huge genetic variation in the genome of *Campylobacter jejuni* is a consequence of intra- genomic process and inherited exchange among various strains. The data from the genome sequences shown variation within these sequences is high; this may be the result of lack of clear homology of many *Campylobacter jejuni* DNA-repair genes. *Campylobacter jejuni* is highly motile by mean of bipolar flagella, play an important role in host colonization [20]. Based on functional and whole genome characterization the composition of flagellum is projected to have sealed composition of hook, basal body, and filament. Filament of flagella composed of two factors, (a) FlaA (b) FlaB, and found involvement of huge amino acid similarity between them [21]. Additionally, *Campylobacter jejuni* flagellar type III secretion system (T3SS) also involved in the secretion of several non-flagellar proteins which play important role in the pathogenesis [21].

Genetic Modifications

The gene which encodes certain enzyme responsible for the biosynthesis of glycan are also found in *Campylobacter jejuni* flagellin and glycotransferase also located near to the flagellin gene is the one of the most hypervariable regions of *Campylobacter* genome [22,23]. The modification in function of glycosylation and its biology is

not fully understood. Conversely polar flagellated bacteria glycosylate the flagellin, mutation in the gene doesn't affect the motility (Logan, 2006). Though all Proteobacteria including *Campylobacter jejuni* , *Campylobacter coli* and *Helicobacter pylori* required glycosylated system to assemble its filament [24]. The *Campylobacter jejuni* flagellin protein carries a terminal sialic acid and most strains of *Campylobacter jejuni* carry gene responsible for the synthesis of two distinct nine carbon sugars: Pseudaminic acid and legionaminic acid acetamidino glycosylation required for the filament assembly [25,26]. Thus, mutation in gene does not affect its motility because the mutant was complemented to encode UDP-N-acetyl-alpha-D-glucosamine C6 dehydratase result in fully motile strains and shows increased solubility as compared to wild type [24]. The region 190 to 310 had substitution of polar amino acid 197 E>D, 202 D>Q, 219 E>D, 251 Q>D and 303 D>E "E" glutamic acid change into "D" aspartic acid, "E" aspartic acid change in to "Q" glutamine or vice versa. Substitution of this amino acid brings change into functioning of gene represented as "#". This study also revealed that the region from 9 to 342 were mutated by isomeric hydrophobic amino acid on position 9 I>V, 157 V>I, 244 I>V, 263 I>V, 275 V>I and 342 I>V or vice versa "I" isoleucine change into "V" valine and marked as "!". The residue number 234 Y>F amino acid change to other isomer of aromatic group of amino acids represented as "%". Similarly inter-genomic recombination or mutation result of horizontal gene transfer recently been discussed between the two mutant of *Campylobacter jejuni* and *Campylobacter coli* strain both contain different resistance to antibiotic marker genes in their flagellin. Our results designate *Campylobacter jejuni* flagellin genes were very well coordinated and predicted residues were conserved with the strain but when compared with the multiple strains its showed huge variability leading toward hypervariable region between multiple strains [27].

Multiple Sequence Alignment

Multiple sequence alignment of Flagellin gene revealed that the hypervariable region of 185 amino acids was found to be conserved over the time while there were three substitutions of similar nature of amino acid were found at 182, 187 and 188 position which brings no change and kept this region partially conserved. Among all the most conserved part with no substitution was from 205-210 and could be the most important part of flagellin protein. The presence of motif region needs to be verified in protein data bank (PDB) [28]. It was also found that evolution in Flagellin *Campylobacter jejuni flaA* and *flaB* genes occurred at the same time. Some regions are conserved among *flaA* and *flaB* of certain isolates while high variability found among other strains within same region. Most of the strains show variability in flagellin genes while the genes present at branch points shows evolution. Moreover, there are selective chunks of sequences differ from *flaA* to *flaB* that are partially conserved among various strains [27].

Current study also revealed in block 245-270, sub block 253 to 258 and 265 to 270 were found to be more conserved within strains. While, there is substitution of hydrophobic amino acid at 263I<V and 243I<V that changes in protein functions are depend on the nature of neighboring amino acids. Phylogenetic analysis of hypervariable sequences enables us to the flagellin genes WP_012006751 was found in *Campylobacter jejuni* strain CJM1cam has ancestor of WP_011812788 genes whereas these both genes were the ancestor of WP_065453915 found in *Campylobacter jejuni* subsp. *jejuni* strain RM1285 and

WP_059110954 was found in 6 strains of *Campylobacter jejuni* sub strains CJ677CC033, CJ677CC034, CJ677CC041, CJ677CC073, CJ677CC085, and CJ677CC539 respectively. Its branch distance also reveals the recombination between its strains is huge. Flagellin genes contain different antibiotic resistance marker and the occurrence of recombination is natural process and the genetic information exchanged between the flagellin genes of *Campylobacter jejuni* strains is through recombination [29].

Evolutionary Phylogenetic Analysis

In present study, we theoretically examined and performed evolutionary phylogenetic analysis based on core and whole genome of closely related 35strains of *Campylobacter jejuni*. In previous study *Campylobacter jejuni* subsp. *Jejuni* strain 01-1512 was the most divergent isolate and showed the presence of Type VI secretion system in gene cluster and the hypervariable regions [30]. *Campylobacter jejuni* subsp. *Jejuni* strain 00-1597 is1742,047 bp contains 1694 genes were similar with the *Campylobacter jejuni* subsp. *Jejuni* strain 01-1512 but strain 00-1597 have different metabolic and virulence properties than other strains [30]. *Campylobacter jejuni* 81-176 have extra 24 set of genes as compared to strain NCTC 11168-Kf1 and NCTC 11168-mfK12E5. Previously reported that the *Campylobacter jejuni* 81-176 have extra 37 genes from the reference strains NCTC 11168 [28] and RM1221 [31] which are located in 11 regions throughout the chromosome [23]. Whereas *Campylobacter jejuni* rest of the strains also showed similarity between maximum numbers of genes. The phylogenetic analysis suggests that the evolution of the *Campylobacter jejuni* strains were acquired by high rate of recombination due to HGT showed variability and diversity in allelic genes which can be varying the disruption of the overall colonal population structures. The large number of polymorphisms is also due to HGT can promptly produce novel phenotypes [23].

CONCLUSION

The *Campylobacter jejuni* was gained huge importance as an infectious agent. Our study increase insight about the genomic variety and possible pathogenic factor of *Campylobacter jejuni* using PGAP tool. In orthologous cluster analysis most of the genes including capsular polysaccharide, flagellar, adhesion and antibiotic resistance genes were found in core genome play important role in virulence. In shared genome most of the hypervariable sequences were found in gene coding flagellin during cluster analysis. Hypervariablity in the flagellin genes lead to disease condition in both humans and animals. In *Campylobacter jejuni* intragenomic recombination and mutation is a result of HGT showed variability and diversity in allelic genes and that is why these bacteria specially *Campylobacter jejuni* contain resistance to different antibiotic genes marker in their flagellin. Flagellin genes considered being an important factor for motility and its role in virulence is more complex than any other method. In Pakistan, Pan-genomics is emerging field and data generated through this study would be extremely useful to understanding the pattern of how's *Campylobacter jejuni* cause illness and showed diversity among evolving strains. This study also supports other scientist to find possible solution for improved treatment and overcome the resistance pattern within the specie. Apart from limited time and resources, in future our focus is to advance current study to do protein modelling and targeted drug development.

REFERENCES

1. Konkel, M. E., Garvis, S. G., Tipton, S. L., Anderson, Jr, D. E., & Cieplak, Jr, W. (1997). Identification and molecular cloning of a gene encoding a fibronectin-binding protein (CadF) from *Campylobacter jejuni*. *Molecular microbiology*, *24*(5), 953- 963.

2. Wiesner, R. S., Hendrixson, D. R., & DiRita, V. J. (2003). Natural transformation of *Campylobacter jejuni* requires components of a type II secretion system. *Journal of bacteriology*, *185*(18), 5408-5418.

3. Mamelli, L., Pagès, J. M., Konkel, M. E., & Bolla, J. M. (2006). Expression and purification of native and truncated forms of CadF, an outer membrane protein of Campylobacter. *International journal of biological macromolecules*, *39*(1-3), 135-140

4.Bozza, C. G., & Pawlowski, W. P. (2008). The cytogenetics of homologous chromosome pairing in meiosis in plants. *Cytogenetic and genome research*, *120*(3-4), 313- 319.

5.Mira, A., Martín-Cuadrado, A. B., D'Auria, G., & Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol*, *13*(2), 45- 57.

6. Rożynek, E., Dzierżanowska-Fangrat, K. A. T. A. R. Z. Y. N. A., Szczepańska, B., Wardak, S., Szych, J., Konieczny, P., ... & Dzierżanowska, D. (2009). Trends in antimicrobial susceptibility of *Campylobacter* isolates in Poland (2000–2007). *Polskie towarzystwo mikrobiologów polish society of microbiologists*, *58*(2), 111-115.

7. Ehrlich, G. D., Hu, F. Z., Shen, K., Stoodley, P., & Post, J. C. (2005). Bacterial plurality as a general mechanism driving persistence in chronic infections. *Clinical orthopaedics and related research*, (437), 20.

8. Georgiades, K., & Raoult, D. (2011). Defining pathogenic bacterial species in the genomic era. *Frontiers in microbiology*, *1*, 151.

9. Thong, K. L., Puthucheary, S. D., & Pang, T. (1997). Genome size variation among recent human isolates of Salmonella typhi. *Research in microbiology*, *148*(3), 229-235.

10. Callister, S. J., McCue, L. A., Turse, J. E., Monroe, M. E., Auberry, K. J., Smith, R. D.,

11. Charlebois, R. L., & Doolittle, W. F. (2004). Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome research*, *14*(12), 2469-2477.

12. Glasner, J. D., Marquez-Villavicencio, M., Kim, H. S., Jahn, C. E., Ma, B., Biehl, B. S., & Dangl, J. L. (2008). Niche-specificity and the variable fraction of the Pectobacterium pan-genome. Molecular plant-microbe interactions, 21(12), 1549-1560.... & Lipton, M. S. (2008). Comparative bacterial proteomics: analysis of the core genome concept. *PloS one*, *3*(2), e1542.

13. Hiller, N. L., Janto, B., Hogg, J. S., Boissy, R., Yu, S., Powell, E., ... & Barbadora, K. (2007). Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. *Journal of bacteriology*, *189*(22), 8186-8195.

14. Lapierre, P., & Gogarten, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends in genetics*, *25*(3), 107-110.

15. McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic acids research*, *32*(suppl_2), W20-W25.

16. King, V., Wassenaar, T., Van Der Zeijst, B. A. M., & Newell, D. G. (1991). Variations in *Campylobacter jejuni* flagellin, and flagellin genes, during in vivo and in vitro passage. *Microbial ecology in health and disease*, *4*(3), 135-140.

17. Ewing, C. P., Andreishcheva, E., & Guerry, P. (2009). Functional characterization of flagellin glycosylation in *Campylobacter jejuni* 81-176. *Journal of bacteriology*, *191*(22), 7086-7093.

18. Fiedoruk, K., Daniluk, T., Rozkiewicz, D., Oldak, E., Prasad, S., & Swiecicka, I. (2019). Whole-genome comparative analysis of *Campylobacter jejuni* strains isolated from patients with diarrhea in northeastern Poland. *Gut pathogens*, *11*(1), 1-10.

19. Wilkinson, D. A., O'Donnell, A. J., Akhter, R. N., Fayaz, A., Mack, H. J., Rogers, L. E., & Midwinter, A. C. (2018). Updating the genomic taxonomy and epidemiology of Campylobacter hyointestinalis. Scientific reports, 8(1), 1-12.

20. Singh, P., & Kwon, Y. M. (2013). Comparative analysis of Campylobacter populations within individual market-age broilers using fla gene typing method. *Poultry science*, *92*(8), 2135-2144.

21. de Zoete, M. R., Keestra, A. M., Wagenaar, J. A., & van Putten, J. P. (2010). Reconstitution of a functional Toll-like receptor 5 binding site in *Campylobacter jejuni* flagellin. *Journal of Biological Chemistry*, *285*(16), 12149-12158.

23. Hofreuter, Kaakoush, N. O., Castaño-Rodríguez, N., Mitchell, H. M., & Man, S. M. (2015). Global epidemiology of Campylobacter infection. *Clinical microbiology reviews*, *28*(3), 687-720.

24. Goon, S., Kelly, J. F., Logan, S. M., Ewing, C. P., & Guerry, P. (2003). Pseudaminic acid, the major modification on Campylobacter flagellin, is synthesized via the Cj1293 gene. *Molecular microbiology*, *50*(2), 659-671.

25. Ketley, J. M. (1997). Pathogenesis of enteric infection by Campylobacter. *Microbiology*, *143*(1), 5-21.

26. Logan, S. M. (2006). Flagellar glycosylation–a new component of the motility repertoire?. *Microbiology*, *152*(5), 1249-1262.

27. Meinersmann, R. J., & Hiett, K. L. (2000). Concerted evolution of duplicate fla genes in Campylobacter. *Microbiology*, *146*(9), 2283-2290.

28. Clark, C. G. (2011). Sequencing of CJIE1 prophages from *Campylobacter jejuni* isolates reveals the presence of inserted and (or) deleted genes. *Canadian journal of microbiology*, *57*(10), 795-809.

29. Wassenaar, T. M., Fry, B. N., & Van der Zeijst, B. A. M. (1995). Variation of the flagellin gene locus of *Campylobacter jejuni* by recombination and horizontal gene transfer. *Microbiology*, *141*(1), 95-101.

30. Clark, C. G., Chen, C. Y., Berry, C., Walker, M., McCorrister, S. J., Chong, P. M., & Westmacott, G. R. (2018). Comparison of genomes and proteomes of four whole genome-sequenced *Campylobacter jejuni* from different phylogenetic backgrounds. *PloS one*, *13*(1), e0190836.

31. Morley, L., McNally, A., Paszkiewicz, K., Corander, J., Méric, G., Sheppard, S. K., ... & Manning, G. (2015). Gene loss and lineage-specific restriction-modification systems associated with niche differentiation in the *Campylobacter jejuni* sequence type 403 clonal complex. *Applied and environmental microbiology*, *81*(11), 3641-3647