Questing functions and structures of hypothetical proteins from *Campylobacter jejuni*: a computer aided approach

Md. Amran Gazi1*, Sultan Mahmud2, Shah Mohammad Fahim1, Md. Rezaul Islam3, Subhasish Das1, Mustafa Mahfuz1,4, Tahmeed Ahmed1

1Nutrition and Clinical Services Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Bangladesh
2Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Bangladesh
3International Max Planck Research School, Grisebachstraße 5, 37077 Göttingen, Germany
4Faculty of Medicine and Life Sciences, University of Tampere, Finland

Sultan Mahmud: sultanmahmud294@yahoo.com
Shah Mohammad Fahim: mohammad.fahim@icddrb.org
Md Rezaul Islam: rezaul.nayeem@gmail.com
Subhasish Das: subhasish.das@icddrb.org
Mustafa Mahfuz: mustafa@icddrb.org
Tahmeed Ahmed: tahmeed@icddrb.org

*Address for correspondence:
Md Amran Gazi
Nutrition and Clinical Services Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Bangladesh, GPO Box 128, Dhaka-1000, Bangladesh.

Email: amran.gazi@icddrb.org
Abstract:

*Campylobacter jejuni* (*C. jejuni*) is considered to be one of the most frequent causes of bacterial gastroenteritis globally, especially in young children. The genome of *C. jejuni* contains many proteins with unknown functions termed as hypothetical proteins (HPs). These proteins might have essential biological role to show the full spectrum of this bacterium. Hence, our study aimed to determine the functions of HPs, pertaining to the genome of *C. jejuni*. An *in silico* work flow integrating various tools were performed for functional assignment, three-dimensional structure determination, domain architecture predictors, sub-cellular localization, physicochemical characterization and protein-protein interactions. Sequences of 267 HPs of *C. jejuni* were analyzed and successfully attributed the function of 49 HPs with higher confidence. Here, we found proteins with enzymatic activity, transporters, binding and regulatory proteins as well as proteins with biotechnological interest. Assessment of the performance of various tools used in this analysis revealed an accuracy of 95% using ROC curve analysis. Functional and structural prediction and the results from ROC analyses provided the validity of *in-silico* tools used in this study. The approach used for this analysis leads us to assign the function of unknown proteins and relate them with the functions that have already been described in previous literature.

**Keywords:** Campylobacter; gastroenteritis; hypothetical protein; *in silico*; NCBI
Introduction

*Campylobacter* is the genus that comprises a diverse group of non-spore forming rod like or spiral shaped Gram-negative bacteria (1). In developing countries, infections with *Campylobacter* are common in children under 2 years of age and found to be associated with increased incidence of diarrheal diseases as well as mortality (1, 2). In industrialized nations, *Campylobacter* is the cause of diarrhea during early years of adulthood (3). *Campylobacter* infections are mostly acquired through consumption of contaminated water and food in resource poor environment (4). Two of the species, *C. jejuni* and *C. coli*, are primarily known to be responsible for human campylobacteriosis (4). Acute gastroenteritis and food poisoning can be induced by *C. jejuni* in infected patients. Usually, *C. jejuni* infection causes gastroenteritis without any complication but acute infection may results in abdominal cramps, fever or other ailments like Guillain-Barré syndrome or Miller Fischer syndrome (5). Recent studies also showed an association of *Campylobacter* infections with malnutrition, a condition highly prevalent in developing countries (2).

Although whole genome sequence of *C. jejuni* NCTC has been published, a detailed catalogue of prospective virulence is yet to be documented. Its complete genome contains a circular chromosome of 1,641,481 base pairs with GC content 30.6%. Several studies since then suggest *C. jejuni* exhibit high genomic diversity across strains. A shotgun DNA microRNA approach revealed 63kb long unique genomic DNA sequences in another *Campylobacter* strain, *C. jejuni* 81-176 when compared to fully sequenced *C. jejuni* NCTC 11168, implying genetic diversity between strains (6, 7). Overall, genome of *C. jejuni* strain 81-176 (total length 1.6 Mb) is available in NCBI encodes 1658 proteins (GC%: 30.4) (7). Among them 267 are yet to be experimentally determined, and are designated as hypothetical proteins (HPs). Similar to functionally annotated proteins, HP originates from an open reading frame (ORF), but lack functional annotations (8). Therefore, annotation of HPs of specific organism leads to the introduction of unique functions, and helps in listing auxiliary protein pathways (8).
Several contemporary bioinformatics tools, for instance, CDART, SMART, Pfam, INTERPROSCAN, MOTIF, SUPERFAMILY and SVMProt has been well established to specify the functions of many bacterial HPs (9-11). Besides, the exploration of protein-protein interaction (PPI) for instance, using STRING database (12), is crucial for comprehending the aspect of biological network. During cellular processes protein interactions plays an essential role. Thus, an understanding of HP function can be reached by studying the PPIs (13). Consequently, interaction of one protein and their function is proven to be dependent on the regulatory connection with other protein [54]. Three-dimensional modeling is also a great way to relate structural knowledge with the function of undetermined proteins (14). Protein structure is generally greater conserved than protein sequence (15). Therefore, structural determination is considered to be a strong indicator of similar function in two or more proteins. Moreover, evolutionary distant proteins and its function can also be identified through structural information (15).

Functional prediction of HPs by using in silico approaches has been successfully applied for various bacteria and parasites (10, 16, 17). In this study, we have chosen C. jejuni as a template to explore the functions of HPs from its genome with a higher accuracy using well optimized bioinformatics tools.

Materials and methods

Retrieval of genome data

Full genome of C. jejuni strain 81-176 was retrieved from NCBI (GCA_000015525.1, NC_008787.1). According to the repository this genome encodes 1658 proteins (http://www.ncbi.nlm.nih.gov/genome/), of which 267 are assigned as HPs. FASTA sequences of HPs were then retrieved for further analysis in this study (accessed February 27, 2019).

Functional analysis of hypothetical proteins

In order to assign the function using the databases depicted in S1 Table, firstly we submitted proteins to five publicly available free tools (CDD-BLAST, HmmScan, SMART, Pfam, and SCANPROSITE) (18-22). These databases can search for the conserved domains and subsequently helps in the categorization
of proteins. Analyses of HPs by five web tools revealed the distinct results. To find a composite result, different confidence levels were assigned on the basis of pooled results obtained from five web-tools. For instance, if we observed same results from the five distinct tools, the composite score was hundred (percentage of confidence). For downstream analyses, we filtered 50 out of 267 HPs that displayed sixty percent or above confidence (Table S2).

Next, we performed functional assignment of these 50 selected HPs using different tools (Fig 1). SMART and CDART (23) facilitated to look for functions using the domain architecture and conserved domain database, respectively. To classify HPs into functional families based on similarity, we employed SUPERFAMILY(24), Pfam (21), and SVMProt (25). Software such as InterPro and MOTIF search tool were also used to detect the motif in the proteins (26, 27). Default parameters were used for all these databases.

We further annotated HPs manually through searching for homologous proteins from related organisms. To do this, we used BLAST against the NCBI nonredundant (nr) database. If the two sequences were ≥ 90% identical, we considered it as homologues to each other. Query cover, score parameters and e-value of every hit are summarized in supplemental material S5.

Geptop 2.0 database was used to identify the essential genes among the HPs (28). Default essentiality score cutoff of 0.24 was adopted. Geptop is the essential gene identification tool based on phylogeny and orthology. In the present study, a similarity search was also done against DrugBank 3.0 for all the targets (29).

**Prediction of physicochemical characteristics**

Expasy's ProtParam server was used for extinction coefficient, isoelectric point (pI), molecular mass, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) prediction (30).
Identification of sub-cellular localization

PSORTb (31) and CELLO (32) were applied to find the localization of HPs in the cell. PSORTb contains the information both from laboratory experimentations and *in silico* prediction. On the other hand, a support vector machine was used by CELLO database to generate the probable localization of protein in the cell. TMHMM (33), SOSUI (34), HMMTOP (35) and SignalP (36) were also applied to detect membrane protein and to verify the presence of cleavage sites for peptide.

Functional protein association networks

We had employed STRING software (37) to predict interactive partners of HPs in this investigation. This database computes the network based on physical and functional associations. Highest score network proteins were selected for this analysis in order to accord the reliability of the PPIs.

Determination of three-dimensional structures

Structure prediction of a protein from its sequences is a way that enables the identification of function. A template based online server PS2-v2 was used to predict the tertiary structure of the HPs in this study (38). This server uses a template of known protein structures and then applied the approaches of multiple and pair-wise alignments combining IMPALA, T-COFFEE and PSI-BLAST.

Performance assessment

A receiver operating characteristic (ROC) was implemented to confirm the accuracy of the predicted functions of HPs from *C. jejuni* genome. Firstly, we selected 40 proteins randomly with known functions of *C. jejuni* (S3 Table). These proteins were predicted for the functions using the same databases that used for the prediction of HPs. To classify the prediction, true positive (1) and true negative (0) were denoted as binary numerals. Six levels diagnostic efficacy was also evaluated where the integers "2", "3", "4" and "5" were used. A web-based calculator was applied to submit the classification data for ROC curve and is
utilized to calculate the sensitivity, specificity, ROC area, and accuracy of the tools used to speculate the function of HPs (39).

Results and discussion

Analysis of HPs from C. jejuni genome

With the ongoing developments of DNA sequencing technologies called high throughput sequencing techniques has enabled a substantial number of bacterial genome sequencing. Annotation of the genes generally depends on sequence homology techniques (40). However, a large number of genes have no assigned function. Therefore, only homology techniques cannot assign functions precisely and may lead to incorrect annotations (41). Multiple tools should be used to avoid this problem to assign functions of HPs. Hence, this study focused on the annotation of HPs from C. jejuni using assorted but effective bioinformatics tools.

Firstly, functional domains were identified from the sequences of all the 267 HPs using SCANPROSITE, SMART, Pfam, CDD-BLAST, and HmmScan. Specific domains could be identified using one, two, three, four or five of the above-stated tools and therefore, different confidence levels were assigned (e.g. 20%, 40%, 60%, 80% and 100%). In our previous studies, published elsewhere, we only considered the proteins with 100% confidence (10, 42). However, in the current study, HPs having 60% or above confidence level has been considered to gain the greater coverage. The analyses revealed 50 such proteins which were used for downstream analyses. For rest of the HPs (n = 217), domains were recognized from one or two of the mentioned tools. Further studies are needed to find the exact function for these proteins. Table S2 summarized protein lists with domain. The final pool of 50 proteins was examined employing CDD-BLAST, Pfam, SMART, MOTIF, InterPro, CDART, SUPERFAMILY, and SVMProt. Functional annotation was considered to be high for proteins that manifested same function from equal or more than three tools (Table S4). Thus, we inferred 49 such proteins with high confidence (Table 1) and classified them as highly confident proteins (Hconf), where 11 contain homologous sequences without product
function reported (S5 Table). Analyses of sequence were then accumulated and Hconf proteins were grouped into different functional categories. Functional classes of proteins consists of regulatory proteins, transporters, binding proteins, enzymes, proteins with biotechnological interest and proteins with other functions (Fig. 2). The categorization was selected based on the literature search and gene ontology. Enzyme classes were determined from enzyme data bank of Expasy (https://enzyme.expasy.org/cgi-bin/enzyme/enzyme-search-cl?).

Moreover, essential genes were predicted using Geptop, a database that accommodates already sequenced bacterial genomes. These genes are fundamental for survival of an organism and perform essential activities of the cell (43). Identification of essential genes is an important stride towards gaining better insight about the evolution (44). Time-absorbing and challenging experiential procedures like transposon mutagenesis, RNA interference, and single-gene knockouts were used to identify essential genes (28). However, in-silico approaches offer an alternative for predicting essential genes. In the current study, it was possible to identify 32 essential proteins by using Geptop database (Table S6). Besides, from the selected Hconf proteins, only one protein was found to be exhibited similarity with approved drugs. The test was done through protein BLAST against DrugBank. Protein WP_002868809.1 showed the similarity with fostamatinib that could act as inhibitors. Drug Bank contains 6816 FDA-approved and experimental drugs, 169 drug enzymes/carriers and 4326 drug targets.

Finally, ROC curve was calculated to identify the reliability of the tools used to predict the function. Average accuracy was found to be 95% for the used pipeline and area under the curve (AUC) was 0.97 (Table 2). It is recommended to use the AUC to summarize the overall accuracy of the tools in the diagnosis (45). The AUC value ranges from 0 to 1, and, the value greater than 0.7 is considered acceptable (45). The ROC analyses results provided the high reliability of in-silico tools used in our study (Table 2). However, predicting the functions of the “function-known” proteins and obtaining very high accuracy does not mean the prediction on “function-unknown” proteins would reproduce the same level of accuracy.
Enzymes

We found five oxidoreductases among these HPs of C. jejuni. These enzymes play key role in the pathogenesis. WP_002824979.1 is NADH-quinone oxidoreductase, an enzyme that involves in regulating the expression of virulence factors, electron transport and sodium translocation (46). This putative domain commonly found in Epsilonproteobacteria, chiefly in Helicobacter pylori (H. pylori) (47). Protein WP_002869225.1 is dimethyl sulphoxide reductase that acts as the terminal electron transfer enzyme in Escherichia coli (E. coli). This enzyme and the reaction it catalyzes could prove helpful on the climate control frontier (48). We also found four proteins as transferase those might involved in bacterial pathogenesis and virulence. Among them protein WP_002854524.1 is responsible for modifying the bacterial character in the presence of repellents and nutrients, found in chemotaxis phosphatase CheX (49). Hydrolases is the third class of enzymes where almost 50% proteins among all characterized enzymes representing this class. This class of proteins is generally membrane bound involved in various virulence factors associated with metal ion binding, transmembrane transport, cell wall degradation. We have found WP_002856630.1 that represents endonuclease-like domain involved in DNA repair and replication (50). WP_009883030.1 and WP_011187235.1 exhibit AAA ATPases (ATPases associated with diverse cellular activities) which plays a number of role in the cell including protein proteolysis and disaggregation, cell-cycle regulation, organelle biogenesis and intracellular transport (51). In addition WP_011187233.1 protein is a toprim (topoisomerase-primase) domain that is found in bacterial DnaG-type primases, involved in DNA strand breakage and rejoining (52).

Binding

We have identified nine proteins as binding among the functionally annotated HPs. These can further classified into RNA binding, DNA binding, protein binding, ion binding and adhesion proteins. Binding of proteins is important in the propagation and survival of pathogens in the host (53). For example, protein binding WP_002868888.1 is tetratricopeptide repeat (TPR) motifs, reported to be directly related to virulence-associated functions (54). WP_002853792.1 is the N-terminal domain of the bacterial
proteins (PgbA) that bind to host cell protein, plasminogen (55). This activity was identified in *H. pylori* where it is thought to contribute to the virulence of this bacterium (55). WP_011117588.1 is mRNA interferase PemK-like domain, a growth inhibitor in *E. coli*. It is responsible for mediating cell death through inhibiting protein synthesis (56). Besides, WP_009882239.1 is haemagglutination activity domain found in a number of large, repetitive proteins of bacteria. Filamentous haemagglutinin (FHA) is a secreted and surface-exposed protein that acts as main virulence attachment factor in childhood whooping cough caused by *Bordetella pertussis* (57). WP_002868809.1 is found to be ankyrin repeat (ANK), a typical protein-protein interaction motif in nature. A large number of bacterial pathogens mimic or manipulate various host functions through delivering ANK-containing proteins into eukaryotic cells (58). Finally, WP_009882608.1 is adhesion protein called surface-exposed lipoprotein JlpA, an early critical step in the pathogenesis of *C. jejuni* disease (59). This HP might provide new approach for the rational design of small molecule inhibitors against *C. jejuni* targeting JlpA efficiently (59).

**Regulatory**

There are six HPs found to be involved in regulatory and cellular mechanisms, and are essential for the pathogenesis of *C. jejuni*, hence can be treated as probable drug targets. WP_002869195.1 is found to be anti-sigma-28 factor that inhibits the activity of the sigma 28 transcription factor. This inhibition prevents the expression of genes from flagellar transcriptional class 3, which include genes for chemotaxis. Mechanism of action of anti-sigma factors has opened new door on the regulation of bacterial gene expression, as anti-sigma factors join another layer to transcriptional control via negative regulation. The bacteriophage T4 uses an anti-sigma factor in order to transcribe its own genes by sabotaging the *E. coli* RNA polymerase (60). WP_002797496.1 is a membrane-associated protein that affects chemotactic events. FliJ is a component of the flagellar export and has a chaperone-like activity. Mutations in FliJ result in failure to respond to chemotactic stimuli (61). Moreover, WP_011117549.1 is identified as conjugal transfer protein that bacteria utilise to export effector molecules during infection. For example, *H. pylori* use type IV machines to transport effectors to the extracellular environment or cell cytosol of
mammalian (62). A DnaA binding protein (WP_002855029.1) HobA, identified that is an essential regulator of DNA replication in *H. pylori* (63). WP_002790076.1 is Methyl-accepting chemotaxis protein (MCP) that allows bacteria to sense the concentrations of molecules (nutrients/toxins) in the extracellular milieu so that they can smooth swim or fall accordingly (64).

**Transporters**

Transporter proteins are involved various metabolic processes, are responsible for transportation of nutrients and hence, essential for survival of the organism. Besides, they accelerate the movement of virulence factors and are directly involved in pathogenesis (65). WP_002855458.1 is the magnesium transporters E (MgtE), found in eukaryotic proteins. Magnesium (Mg2+) is an essential element for growth and maintenance of living cells where MgtE transport magnesium across the cell membrane (66).

WP_002868880.1 is ABC-type transport, responsible for outer membrane biosynthesis in bacteria that can be an excellent drug targets (67). WP_002856180.1 is HMA domain (heavy-metal-associated domain) found in a number of detoxification proteins or in heavy metals transport. Proteins that involved in transporting heavy metals in bacteria, plant and mammals share similarities across the kingdoms in their structures and sequences. These proteins provide an important arena for research, some being involved in bacterial resistance to toxic metals, while others are responsible for acquired human diseases, such as Wilson's and Menke's diseases (68). WP_011117548.1 is the bacterial virulence protein VirB8 that is thought to be a constituent of DNA transporter. In addition, VirB8 is a potential drug target that targets its protein-protein interactions. X-ray structure has enabled a detailed structure-function analysis of VirB8, which identified VirB8 interaction with VirB4 and VirB10 (69). Our results also go in line with this as we observed VirB8 has strong interaction with VirB10.

**Potential proteins with biotechnological application**

We identified few proteins that can have biotechnological applications based on their functional process. For instance, WP_010790856.1 is pyridoxamine 5'-phosphate oxidase (pdxH), an enzyme involved in the
de novo synthesis of pyridoxal phosphate and pyridoxine (vitamin B6). Moreover, PdxH is evolutionary related to phzD (also known as phzG), one of the enzymes in the phenazine biosynthesis protein pathway (70). Only known source of phenazines are bacteria in nature. This is used as drug and also acts as biocounter agents to inhibit plant pests. For example, the phenazine pyocyanin contributes to its potential to colonise the lungs of cystic fibrosis patients (71). Similarly, phenazine-1-carboxylic acid, produced by a number of Pseudomonas, increases survival in soil and has been shown to be important for the biological control of certain strains (72). The protein WP_002869072.1 was predicted to be S-adenosyl-L-methionine-dependent methyltransferase (SAM-MTase). Methyltransferases transfer a methyl group from a donor to an acceptor during methylation of biopolymers (73). SAM-MT was used in the pharmaceutical industry as catechol, first as an anti-microbial and anti-cancer agent (73, 74).

Protein WP_024088174.1 is the nitrate reductases that produce nitrite from nitrate. Nitrate is the primary source of nitrogen in fertilized soils and the reaction is critical for the production of protein in crop plants. Nitrate reductase enzyme activity can also be used as a biochemical tool for predicting grain protein production and subsequent grain yield. For example, it promotes amino acid content in tea leaves (75). It is also reported that tea plants sprayed with various micronutrients (like Zn, Mn and B) along with Mo enhanced the amino acid production of tea and the crop yield (75). WP_002869028.1 is a phytase like domain that catalyzes the hydrolysis of phytic acid. Phytic acid is organic form of phosphorus and indigestible found in grains and oil seeds. Phytase is produced by bacteria found in the gut of ruminant animals is able to make phosphorus from phytic acid (76). But, non-ruminants like human cannot make phytase. Research in the field of animal nutrition has put the idea of supplementing feed with phytase to make sure the availability of phytate-bound nutrients like phosphorus, calcium, carbohydrates, proteins and other minerals (77).

Peptidases, an enzyme that is used as the ingredients of detergents, foods and pharmaceuticals (78). In this study, WP_009882583.1 was found to be cysteine peptidase that hydrolyses a peptide bond utilizing the thiol group of cysteine as nucleophile. These peptidases are often confined to acidic environments and
active at acidic pH such as the plant vacuole or animal lysosome. WP_002868905.1 is GDSL esterases and lipases are hydrolytic enzymes with broad substrate specificity. They have potential for use in the synthesis and hydrolysis of ester compounds of biochemical, food, pharmaceutical, and other biological interests (79).

**Other proteins**

WP_002856369.1 and WP_002856602.1 was found to be beta-lactamase-inhibitor, a group of enzymes responsible for bacterial resistance to beta-lactam antibiotics (80). WP_009883121.1 act as Flagellar FliS export co-chaperone. Previously, various FliS-associated proteins in *H. pylori* were identified by a yeast 2-hybrid study, but the implications are unknown (81). Chaperons are usually involved in various important processes such as protein degradation, folding, and polypeptide translocation (81).

Lastly, WP_002860117.1 protein family includes two enzymes involved in menaquinone (vitamin K2) biosynthesis. In prokaryotes, vitamin K2 serves as the sole quinone molecule in electron shuffling systems while menaquinone pathway is absent in humans (82). Therefore, novel antibacterial agents are possible to develop by targeting the bacterial enzymes responsible for menaquinone biosynthesis. It has been reported that inhibition of menaquinone showed significant growth inhibition against multidrug-resistant *Mycobacterium* and other gram-positive bacteria as well as effective in killing gram-negative bacteria (83).

**Prediction of primary properties and protein localization**

Sequences of amino acids of 49 HPs were analyzed to evaluate their primary properties, and their localization (Table S7). But, we paid attention to some proteins that showed functions important for the survival of Campylobacter and might have biotechnological interest. The proteins WP_024088174.1, WP_002869072.1, WP_010790856.1, WP_002868905.1, WP_002869028.1, WP_009882583.1 all had molecular weight (MW) values between 15792.47 and 52423.83. These proteins are referred as biotechnologically important in this study. Some proteins, essential for pathogenesis of Campylobacter
have MW ranged from 8773.25 to 39113.6. The isoelectric point is the pH where protein carries no net electrical charge. For the list of mentioned proteins, it ranged from 5.03 to 9.63.

The aliphatic index indicates the protein thermo stability (84). Protein WP_002856369.1, associated with beta-lactamase inhibition showed the highest values of 133.14. The grand average of hydropathy (GRAVY) of protein indicates its hydrophobicity or the interaction with water (85). In WP_002869028.1, WP_009882583.1 and WP_02408174.1, the scores are between -0.744, -0.439 and -0.393. Moreover, the instability index offers an assumption of the stability of protein in vitro. We used cutoff values >40 and <40 to discriminate between stable and unstable proteins respectively. From our listed proteins, WP_02408174.1 and WP_002868880.1 were considered to be stable.

Localization plays an essential role in determining function of unknown proteins (11). Protein WP_002868905.1 and WP_009882583.1 is located in outer membrane whereas other proteins of interest were predicted to be in the cytoplasm.

**Protein-protein interaction network**

Function of a completely unknown protein can be identified based on the evidence of their interactions with the known proteins of a particular organism (11). For example, PPI map and *in-vitro* proteome-wide interaction screens were applied to successfully assign the function of fifty unknown proteins for *Streptococcus pneumonia* (86). In our study protein WP_010790856.1, an oxidase (pdxH) showed a strong interaction with the Pyridoxine 5’-phosphate synthase that involved in vitamin B6 synthesis. WP_02408174.1 is interacted with formate dehydrogenase, an oxidoreductase that oxidizes formate to form carbon dioxide. WP_002868880.1 was found to be interacted with ABC transporter that functions to maintain the asymmetry of the outer membrane. All these predictions of functional partners have strengthened our findings of function predicted by using functional prediction tools (S8 Table).
Three-dimensional structures

Structural genomics has become a robust way to determine the novel structures of proteins, especially via X-ray crystallography (87). Determination of unannotated protein structures can often help us to discover unexpected family relationships, hence giving the idea of their probable functions. Proteins unrelated to existing PDB entries may represent new functions. In this case structures homologous to other organism have manifested as surrogates in drug discovery. For example, Nolatrexed an anti-cancer drug was discovered using the structure of E. coli thymidylate synthase (46% sequence identity with human homolog) (87). Kinase inhibitors to kill the Plasmodium falciparum were identified using structures of protein kinases from Cryptosporidium and Toxoplasma (61% and 74% sequence identity respectively) (88).

In our study, PS2-v2 online server was used to model the three dimensional structures of the Hconf proteins for Campylobacter. Among the 49 Hconf proteins, 24 proteins revealed same domain as function prediction tools used in this study. In contrast, 9 proteins showed discrepant results and no suitable templates were found for 16 proteins (Table S9). Identity of model was ranging from 54.5% to 91.6% and was constructed from closely related Campylobacter genus bacteria belonging to the H. pylori, E. coli, Bacillus and Clostridium.

Based on the resolution and identity, two best models were WP_002797496.1 and WP_002854991.1, which were annotated as Flagellar FliJ protein and FxsA cytoplasmic membrane protein respectively. The structure obtained for FliJ protein was determined by X-ray crystallography earlier and refined with diffraction data to 1.8 Å resolutions, which was solved by an ortholog isolated from Saccharomyces cerevisiae (PDB 2efrA). FxsA was determined by electron microscopy and refined with diffraction data to 4 Å resolutions and solved by an ortholog isolated from Torpedo marmorata (PDB 1oedB). Both of these proteins showed the same function as predicted by other function prediction tools. Proteins with shared sequence typically display similar functions in this way.
Conclusions

Protein function identification of a pathogen is an essential step to understand its cellular and molecular processes. In this study, we used a computer aided approach to assign the function of HPs from *C. jejuni*. We predicted the function to 49 HPs with a higher confidence. In addition, localization of protein and primary structure prediction were useful in supporting the specific characteristics of annotated proteins. Proteins were further explored for PPI and their tertiary structures. We have identified proteins with important functions including enzymes, transporters, binding and regulatory proteins as well as proteins with biotechnological interest. To summarize, our comprehensive analysis produces a better understanding of *C. jejuni* genome related HPs that would help to find novel therapeutic interventions and targets. Moreover, we have obtained an excellent result using the pipeline used in this study and the method can be used to annotate the function of unknown proteins.

However, biochemical and clinical investigations are required to confirm the function of predicted proteins. Several studies have been conducted previously using the cumulative *in-silico* and *in-vitro/in-vivo* approach to investigate the function of unknown proteins. For instance, *in silico* approaches were used to predict the biological function of some of the unknown *Mycobacterium* proteins. The chosen proteins posses the α/β-hydrolase topological fold, characteristic of lipases/esterases which were further validated by wet lab experiments (89). Combination of *in-silico* and *in-vitro/in-vivo* assays were also used to characterize the function of HPs from several other organisms (90-93). Moreover, *in-silico* structure prediction methods were applied for drug discovery in the absence of x-ray structure of the target protein and again confirmed by *in-vitro* assays. Nonetheless, functional prediction merely on *in silico* methods requires careful integration of several computational tools into a single streamlined process. We hope that the information of HPs in this study will be innovative for further *in-vitro/in-vivo* analysis on *C. jejuni*. 
Author's contributions

MAG has made substantial contributions to conception, design and drafting the manuscript. SM, SMF, MRI, SD participated in the acquisition, analysis and interpretation of data. MM and TA conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to core donors which provide unrestricted support to icddr,b for its operations and research. Current donors providing unrestricted support include: Government of the People’s Republic of Bangladesh; Canadian International Development Agency (CIDA), Swedish International Development Cooperation Agency (Sida), and the Department for International Development, UK (DFID). We gratefully acknowledge these donors for their support and commitment to icddr,b’s research efforts.

Competing interests

The authors declare that they have no competing interests

Funding

No funding declared

References

1. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. Global epidemiology of Campylobacter infection. Clinical microbiology reviews. 2015;28(3):687-720.

2. Platts-Mills JA, Kosek M. Update on the burden of Campylobacter in developing countries. Current opinion in infectious diseases. 2014;27(5):444.
3. Mehla K, Ramana J. Novel drug targets for food-borne pathogen Campylobacter jejuni: an integrated subtractive genomics and comparative metabolic pathway study. Omics: a journal of integrative biology. 2015;19(7):393-406.

4. Coker AO, Isokpehi RD, Thomas BN, Amisu KO, Obi CL. Human campylobacteriosis in developing countries. Emerging infectious diseases. 2002;8(3):237.

5. Takahashi M, Koga M, Yokoyama K, Yuki N. Epidemiology of Campylobacter jejuni isolated from patients with Guillain-Barré and Fisher syndromes in Japan. Journal of clinical microbiology. 2005;43(1):335-9.

6. Poly F, Threadgill D, Stintzi A. Genomic diversity in Campylobacter jejuni: identification of C. jejuni 81-176-specific genes. Journal of clinical microbiology. 2005;43(5):2330-8.

7. Parkhill J, Wren B, Mungall K, Ketley J, Churcher C, Basham D, et al. The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. Nature. 2000;403(6770):665.

8. Nimrod G, Schushan M, Steinberg DM, Ben-Tal N. Detection of functionally important regions in “hypothetical proteins” of known structure. Structure. 2008;16(12):1755-63.

9. Shahbaaz M, ImtaiyazHassan M, Ahmad F. Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. PloS one. 2013;8(12):e84263.

10. Gazi MA, Kibria MG, Mahfuz M, Islam MR, Ghosh P, Afsar MNA, et al. Functional, structural and epitopic prediction of hypothetical proteins of Mycobacterium tuberculosis H37Rv: an in silico approach for prioritizing the targets. Gene. 2016;591(2):442-55.

11. da Costa WLO, de Aragão Araújo CL, Dias LM, de Sousa Pereira LC, Alves JTC, Araújo FA, et al. Functional annotation of hypothetical proteins from the Exiguobacterium antarcticum strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PloS one. 2018;13(6):e0198965.
12. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic acids research. 2016:gkw937.

13. Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I. Fundamentals of protein interaction network mapping. Molecular systems biology. 2015;11(12):848.

14. Jez JM. Revisiting protein structure, function, and evolution in the genomic era. Journal of invertebrate pathology. 2017;142:11-5.

15. Gherardini PF, Helmer-Citterich M. Structure-based function prediction: approaches and applications. Briefings in Functional Genomics and Proteomics. 2008;7(4):291-302.

16. Varma PBS, Adimulam YB, Kodukula S. In silico functional annotation of a hypothetical protein from Staphylococcus aureus. Journal of infection and public health. 2015;8(6):526-32.

17. Ravooru N, Ganji S, Sathyarayanan N, Nagendra HG. Insilico analysis of hypothetical proteins unveils putative metabolic pathways and essential genes in Leishmania donovani. Frontiers in genetics. 2014;5:291.

18. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic acids research. 2014;43(D1):D222-D6.

19. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic acids research. 2011;39(suppl_2):W29-W37.

20. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. Nucleic acids research. 2000;28(1):231-4.

21. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic acids research. 2013;42(D1):D222-D30.

22. De Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic acids research. 2006;34(suppl_2):W362-W5.
23. Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome research. 2002;12(10):1619-23.

24. Wilson D, Madera M, Vogel C, Chothia C, Gough J. The SUPERFAMILY database in 2007: families and functions. Nucleic acids research. 2006;35(suppl_1):D308-D13.

25. Cai C, Han L, Ji ZL, Chen X, Chen YZ. SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic acids research. 2003;31(13):3692-7.

26. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic acids research. 2009;37(suppl_2):W202-W8.

27. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017—beyond protein family and domain annotations. Nucleic acids research. 2016;45(D1):D190-D9.

28. Wei W, Ning L-W, Ye Y-N, Guo F-B. Geptop: a gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. PloS one. 2013;8(8):e72343.

29. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for ‘omics’ research on drugs. Nucleic acids research. 2010;39(suppl_1):D1035-D41.

30. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook: Springer; 2005. p. 571-607.

31. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010;26(13):1608-15.

32. Yu CS, Lin CJ, Hwang JK. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein science. 2004;13(5):1402-6.
33. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology. 2001;305(3):567-80.

34. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics (Oxford, England). 1998;14(4):378-9.

35. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics. 2001;17(9):849-50.

36. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature biotechnology. 2019;1.

37. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic acids research. 2014;43(D1):D447-D52.

38. Chen C-C, Hwang J-K, Yang J-M. 2-v2: template-based protein structure prediction server. Bmc Bioinformatics. 2009;10(1):366.

39. Eng J. ROC analysis: web-based calculator for ROC curves. Baltimore: Johns Hopkins University [updated 2014 March 19]. Accessed 25/05/17. Available from: http://www.jrocfit.org; 2017.

40. Pearson WR. An introduction to sequence similarity (“homology”) searching. Current protocols in bioinformatics. 2013;42(1):3.1. -3.1. 8.

41. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS computational biology. 2009;5(12):e1000605.

42. Gazi MA, Mahmud S, Fahim SM, Kibria MG, Palit P, Islam MR, et al. Functional Prediction of Hypothetical Proteins from Shigella flexneri and Validation of the Predicted Models by Using ROC Curve Analysis. Genomics & informatics. 2018;16(4).

43. Zhang Z, Ren Q. Why are essential genes essential?-The essentiality of Saccharomyces genes. Microbial Cell. 2015;2(8):280.
44. Romero IG, Ruvinsky I, Gilad Y. Comparative studies of gene expression and the evolution of gene regulation. Nature Reviews Genetics. 2012;13(7):505.

45. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian journal of internal medicine. 2013;4(2):627.

46. Verkhovsky MI, Bogachev AV. Sodium-translocating NADH: quinone oxidoreductase as a redox-driven ion pump. Biochimica et Biophysica Acta (BBA)-Bioenergetics. 2010;1797(6-7):738-46.

47. Wang G, Maier RJ. An NADPH quinone reductase of Helicobacter pylori plays an important role in oxidative stress resistance and host colonization. Infection and immunity. 2004;72(3):1391-6.

48. Kroneck PM, Torres MES. The metal-driven biogeochemistry of gaseous compounds in the environment: Springer; 2014.

49. Simon MI, Borkovich KA, Bourret RB, Hess JF. Protein phosphorylation in the bacterial chemotaxis system. Biochimie. 1989;71(9-10):1013-9.

50. Hosfield DJ, Mol CD, Shen B, Tainer JA. Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. Cell. 1998;95(1):135-46.

51. Kedzierska S. Structure, function and mechanisms of action of ATPases from the AAA superfamily of proteins. Postepy biochemii. 2006;52(3):330-8.

52. Aravind L, Leipe DD, Koonin EV. Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic acids research. 1998;26(18):4205-13.

53. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Introduction to pathogens. Molecular Biology of the Cell 4th edition: Garland Science; 2002.

54. Cerveny L, Straskova A, Dankova V, Hartlova A, Ceckova M, Staud F, et al. Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. Infection and immunity. 2013;81(3):629-35.
55. Jönsson K, Guo BP, Monstein H-J, Mekalanos JJ, Kronvall G. Molecular cloning and characterization of two Helicobacter pylori genes coding for plasminogen-binding proteins. Proceedings of the National Academy of Sciences. 2004;101(7):1852-7.

56. Zhang J, Zhang Y, Zhu L, Suzuki M, Inouye M. Interference of mRNA function by sequence-specific endoribonuclease PemK. Journal of Biological Chemistry. 2004;279(20):20678-84.

57. Makhov A, Hannah J, Brennan M, Trus B, Kocsis E, Conway J, et al. Filamentous hemagglutinin of Bordetella pertussis: a bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. Journal of molecular biology. 1994;241(1):110-24.

58. Al-Khodor S, Price CT, Kalia A, Kwaik YA. Functional diversity of ankyrin repeats in microbial proteins. Trends in microbiology. 2010;18(3):132-9.

59. Kawai F, Paek S, Choi K-J, Prouty M, Kanipes MI, Guerry P, et al. Crystal structure of JlpA, a surface-exposed lipoprotein adhesin of Campylobacter jejuni. Journal of structural biology. 2012;177(2):583-8.

60. Tlapák H, Rydzewski K, Schulz T, Weschka D, Schunder E, Heuner K. Functional analysis of the alternative sigma-28 factor FliA and its anti-sigma factor FlgM of the nonflagellated Legionella species L. oakridgensis. Journal of bacteriology. 2017;199(11):e00018-17.

61. Minamino T, Chu R, Yamaguchi S, Macnab RM. Role of FliJ in Flagellar Protein Export in Salmonella. Journal of Bacteriology. 2000;182(15):4207-15.

62. Christie PJ, Vogel JP. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. Trends in microbiology. 2000;8(8):354-60.

63. Natraj G, Noirot-Gros MF, Zawilak-Pawlik A, Kapp U, Terradot L. The structure of a DnaA/HobA complex from Helicobacter pylori provides insight into regulation of DNA replication in bacteria. Proceedings of the National Academy of Sciences. 2009;106(50):21115-20.

64. Ud-Din AIMS, Roujeinikova A. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. Cellular and molecular life sciences. 2017;74(18):3293-303.
65. Yuan J, Zweers JC, Van Dijl JM, Dabey RE. Protein transport across and into cell membranes in bacteria and archaea. Cellular and Molecular Life Sciences. 2010;67(2):179-99.

66. Yan Y-W, Mao D-D, Yang L, Qi J-L, Zhang X-X, Tang Q-L, et al. Magnesium transporter MGT6 plays an essential role in maintaining magnesium homeostasis and regulating high magnesium tolerance in Arabidopsis. Frontiers in plant science. 2018;9:274.

67. Bugde P, Biswas R, Merien F, Lu J, Liu D-X, Chen M, et al. The therapeutic potential of targeting ABC transporters to combat multi-drug resistance. Expert opinion on therapeutic targets. 2017;21(5):511-30.

68. Bull PC, Cox DW. Wilson disease and Menkes disease: new handles on heavy-metal transport. Trends in Genetics. 1994;10(7):246-52.

69. Bailey S, Ward D, Middleton R, Grossmann JG, Zambryski PC. Agrobacterium tumefaciens VirB8 structure reveals potential protein–protein interaction sites. Proceedings of the National Academy of Sciences. 2006;103(8):2582-7.

70. Pierson III LS, Gaffney T, Lam S, Gong F. Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium Pseudomonas aureofaciens 30-84. FEMS microbiology letters. 1995;134(2-3):299-307.

71. Hunter RC, Klepac-Ceraj V, Lorenzi MM, Grotzinger H, Martin TR, Newman DK. Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung function and microbial complexity. American journal of respiratory cell and molecular biology. 2012;47(6):738-45.

72. Upadhyay A, Srivastava S. Phenazine-1-carboxylic acid is a more important contributor to biocontrol Fusarium oxysporum than pyrrolnitrin in Pseudomonas fluorescens strain Psd. Microbiological research. 2011;166(4):323-35.

73. Banco MT, Mishra V, Greeley SC, Ronning DR. Direct Detection of Products from S-Adenosylmethionine-Dependent Enzymes Using a Competitive Fluorescence Polarization Assay. Analytical chemistry. 2018;90(3):1740-7.
74. Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Current opinion in structural biology. 2002;12(6):783-93.

75. Ruan J, Wu X, Ye Y, Härdter R. Effect of potassium, magnesium and sulphur applied in different forms of fertilisers on free amino acid content in leaves of tea (Camellia sinensisL). Journal of the Science of Food and Agriculture. 1998;76(3):389-96.

76. Frias J, Doblado R, Antezana JR, Vidal-Valverde C. Inositol phosphate degradation by the action of phytase enzyme in legume seeds. Food Chemistry. 2003;81(2):233-9.

77. Dersjant-Li Y, Awati A, Schulze H, Partridge G. Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. Journal of the Science of Food and Agriculture. 2015;95(5):878-96.

78. Harada J, Takaku S, Watanabe K. An on-demand metalloprotease from psychro-tolerant Exiguobacterium undae Su-1, the activity and stability of which are controlled by the Ca2+ concentration. Bioscience, biotechnology, and biochemistry. 2012;110997.

79. Akoh CC, Lee G-C, Liaw Y-C, Huang T-H, Shaw J-F. GDSL family of serine esterases/lipases. Progress in lipid research. 2004;43(6):534-52.

80. KONG KF, Schneper L, Mathee K. Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. Apmis. 2010;118(1):1-36.

81. Lam WWL, Woo EJ, Kotaka M, Tam WK, Leung YC, Ling TKW, et al. Molecular interaction of flagellar export chaperone FliS and cochaperone HP1076 in Helicobacter pylori. The FASEB Journal. 2010;24(10):4020-32.

82. Meganathan R. Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms. Vitamins & Hormones. 2001;61:173-218.

83. Debnath J, Siricilla S, Wan B, Crick DC, Lenaerts AJ, Franzblau SG, et al. Discovery of selective menaquinone biosynthesis inhibitors against Mycobacterium tuberculosis. Journal of medicinal chemistry. 2012;55(8):3739-55.
84. Idicula-Thomas S, Balaji PV. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. Protein Science. 2005;14(3):582-92.

85. Jaspard E, Macherel D, Hunault G. Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. PLoS One. 2012;7(5):e36968.

86. Meier M, Sit RV, Quake SR. Proteome-wide protein interaction measurements of bacterial proteins of unknown function. Proceedings of the National Academy of Sciences. 2013;110(2):477-82.

87. Chance MR, Bresnick AR, Burley SK, Jiang JS, Lima CD, Sali A, et al. Structural genomics: a pipeline for providing structures for the biologist. Protein Science. 2002;11(4):723-38.

88. Cardew EM, Verlinde CL, Pohl E. The calcium-dependent protein kinase 1 from Toxoplasma gondii as target for structure-based drug design. Parasitology. 2018;145(2):210-8.

89. Kumar A, Sharma A, Kaur G, Makkar P, Kaur J. Functional characterization of hypothetical proteins of Mycobacterium tuberculosis with possible esterase/lipase signature: a cumulative in silico and in vitro approach. Journal of Biomolecular Structure and Dynamics. 2017;35(6):1226-43.

90. Choi H-P, Juarez S, Ciordia S, Fernandez M, Bargiela R, Albar JP, et al. Biochemical characterization of hypothetical proteins from Helicobacter pylori. PLoS One. 2013;8(6).

91. Cort JR, Yee A, Edwards AM, Arrowsmith CH, Kennedy MA. NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from Methanobacterium thermoautotrophicum. Journal of structural and functional genomics. 2000;1(1):15-25.

92. Barta ML, Thomas K, Yuan H, Lovell S, Batalla KP, Schramm VL, et al. Structural and biochemical characterization of Chlamydia trachomatis hypothetical protein CT263 supports that menaquinone synthesis occurs through the futalosine pathway. Journal of Biological Chemistry. 2014;289(46):32214-29.
93. Zhang W, Culley DE, Gritsenko MA, Moore RJ, Nie L, Scholten JC, et al. LC–MS/MS based proteomic analysis and functional inference of hypothetical proteins in Desulfovibrio vulgaris. Biochemical and biophysical research communications. 2006;349(4):1412-9.

Figure Legends:

Fig. 1. Flow chart showing the overall design of the study

Fig. 2. Functional classification of 49 HPs into various groups

Table 1. HPs functionally annotated from *C. jejuni*.

| No. | Protein IDs     | Protein function                                              |
|-----|----------------|---------------------------------------------------------------|
| 1   | WP_002868767.1 | Curli production assembly, transport component CsgG          |
| 2   | WP_002854524.1 | Chemotaxis phosphatase CheX                                  |
| 3   | WP_009882162.1 | SprA-related family                                           |
| 4   | WP_010790856.1 | Pyridoxamine 5'-phosphate oxidase                            |
| 5   | WP_009882239.1 | Haemagglutination activity domain                            |
| 6   | WP_002854991.1 | FxsA cytoplasmic membrane protein, FxsA                       |
| 7   | WP_002855029.1 | DNA replication regulator, HobA                              |
| 8   | WP_002868905.1 | GDSL-like Lipase                                              |
| 9   | WP_002869356.1 | Divergent polysaccharide deacetylase                         |
| 10  | WP_002856929.1 | C4-type zinc ribbon domain                                    |
| 11  | WP_002869028.1 | Esterase-like activity of phytase                             |
| 12  | WP_011812736.1 | Domain of unknown function DUF234                            |
| 13  | WP_002868809.1 | Ankyrin repeats, Ank_2                                       |
| 14  | WP_002869368.1 | Type-1V conjugative transfer system mating pair stabilisation, TraN |
| 15  | WP_009882583.1 | NLPC_P60 stabilising domain                                  |
| 16  | WP_002853389.1 | Jag, N-terminal domain superfamily                           |
| 17  | WP_009882608.1 | Adhesin from Campylobacter                                   |
| 18  | WP_002856369.1 | Putative beta-lactamase-inhibitor-like                        |
| No. | Accession | Description |
|-----|-----------|-------------|
| 19  | WP_079254190.1 | Beta-1,4-N-acetylglactosaminytransferase (CgtA) |
| 20  | WP_002856180.1 | Heavy-metal-associated domain |
| 21  | WP_002831611.1 | Transcription factor zinc-finger |
| 22  | WP_002790076.1 | Methyl-accepting chemotaxis protein (MCP) signalling domain |
| 23  | WP_002853792.1 | Plasminogen-binding protein pgbA N-terminal |
| 24  | WP_002869072.1 | Putative S-adenosyl-L-methionine-dependent methyltransferase |
| 25  | WP_002869097.1 | MaoC-like dehydratase domain |
| 26  | WP_002869326.1 | Metallo-carboxypeptidase |
| 27  | WP_002869139.1 | Pyruvate phosphate dikinase, PEP |
| 28  | WP_002869195.1 | Anti-sigma-28 factor |
| 29  | WP_002856630.1 | PD-(D/E)XK nuclease superfamily |
| 30  | WP_002855458.1 | MgtE intracellular N domain |
| 31  | WP_002797496.1 | Flagellar FliJ protein |
| 32  | WP_024088174.1 | Nitrate reductase chaperone |
| 33  | WP_009883030.1 | ATPase, AAA-type, core |
| 34  | WP_002824979.1 | putative NADH-ubiquinone oxidoreductase chain E |
| 35  | WP_002869225.1 | DMSO reductase anchor subunit (DmsC) |
| 36  | WP_002856602.1 | Putative beta-lactamase-inhibitor-like |
| 37  | WP_002868888.1 | Tetra-tricopeptide repeat, TPR_2 |
| 38  | WP_002868880.1 | ABC-type transport auxiliary lipoprotein component |
| 39  | WP_009883121.1 | Flagellar FliS export co-chaperone |
| 40  | WP_002860117.1 | Menaquinone biosynthesis |
| 41  | WP_002779704.1 | T-antigen specific domain |
| 42  | WP_011187233.1 | Toprim domain |
| 43  | WP_011187235.1 | AAA domain, AAA_25 |
| 44  | WP_002809111.1 | TrbM Superfamily |
| 45  | WP_011117548.1 | Bacterial virulence protein VirB8 |
| 46  | WP_011117549.1 | Conjugal transfer protein |
| 47  | WP_011117575.1 | Type IV secretion system proteins,T4SS |
| 48  | WP_011799393.1 | TrbM Superfamily |
| 49  | WP_011117588.1 | mRNA interferase PemK-like |
Table 2. ROC results of various tools used in this study

| No. | Software      | Accuracy (%) | Sensitivity (%) | Specificity (%) | ROC area |
|-----|---------------|--------------|-----------------|-----------------|----------|
| 1   | PFAM          | 95%          | 94.7%           | 100%            | 0.97     |
| 2   | SMART         | 95%          | 94.9%           | 100%            | 0.97     |
| 3   | MOTIF         | 95%          | 94.9%           | 100%            | 0.97     |
| 4   | INTERPROSCAN  | 95%          | 94.9%           | 100%            | 0.97     |
| 5   | CDART         | 97.5%        | 97.4%           | 100%            | 0.99     |
| 6   | SUPERFAMILY   | 95%          | 94.1%           | 100%            | 0.97     |
| 7   | SVMprot       | 90%          | 88.9%           | 100%            | 0.94     |
| 8   | Average       | 95%          | 94.3%           | 100%            | 0.97     |
Full protein profile retrieved *Campylobacter jejuni* (1658)

Filter hypothetical proteins (267)

Domain identification by five different domain databases

Domain identified by three or more databases (60% or above confidence)

Yes (50)

Functional prediction: pfam, SMART, MOTIF, InterPro, CDART, SUPERFAMILY, SVMprot

HP with function predicted by three or more tools

Yes (49)

Manual curation by search of homologous sequences: Blast

Search for essential genes: Geptop

Characterization: Psortb, CELLO, signalP, HMMTOP, TMHMM, SOSUI, ProtParam, STRING, PS2-V2

Inference of protein function
## S1 Table. List of bioinformatics tools and databases

| Analyse                        | Bioinformatics tool | Version | URL                                      |
|-------------------------------|---------------------|---------|------------------------------------------|
| **Functional analysis and conserved domain** |                      |         |                                          |
|                                | Pfam                | 31.0    | [https://pfam.xfam.org/](https://pfam.xfam.org/) |
|                                | SMART               | 8.0     | [http://smart.embl-heidelberg.de/](http://smart.embl-heidelberg.de/) |
|                                | MOTIF               | *       | [https://www.genome.jp/tools/motif/](https://www.genome.jp/tools/motif/) |
|                                | InterPro            | 66.0    | [https://www.ebi.ac.uk/interpro/](https://www.ebi.ac.uk/interpro/) |
|                                | CDART               | *       | [https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi](https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi) |
|                                | SUPERFAMILY         | 1.75    | [http://supfam.org/SUPERFAMILY/index.html](http://supfam.org/SUPERFAMILY/index.html) |
|                                | SVMProt             | *       | [http://bidd2.nus.edu.sg/cgi-bin/svmprot/svmprot.cgi](http://bidd2.nus.edu.sg/cgi-bin/svmprot/svmprot.cgi) |
|                                | CDD-Blast           | 3.16    | [https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi](https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) |
|                                | HmmScan             | 3.2.1   | [https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan](https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan) |
|                                | Scanprosite         | *       | [https://prosite.expasy.org/scanprosite/](https://prosite.expasy.org/scanprosite/) |
|                                | Geptop              | 2.0     | [http://cefg.uestc.cn/geptop/](http://cefg.uestc.cn/geptop/) |
| **Sub-cellular localization of the protein** |                      |         |                                          |
|                                | PSORTdb             | 3.0     | [http://db.psort.org/](http://db.psort.org/) |
|                                | CELLO               | 2.5     | [http://cello.life.nctu.edu.tw/](http://cello.life.nctu.edu.tw/) |
|                                | SignalP             | 5.0     | [http://www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/) |
|                                | HMMPROTEIN          | *       | [http://www.enzim.hu/hmmtop/](http://www.enzim.hu/hmmtop/) |
|                                | TMHMMP              | 2.0     | [http://www.cbs.dtu.dk/services/TMHMM/](http://www.cbs.dtu.dk/services/TMHMM/) |
|                                | SOSUI               | *       | [http://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html](http://harrier.nagahama-i-bio.ac.jp/sosui/sosui_submit.html) |
| **Physical-chemical characterization** |                      |         |                                          |
|                                | ProtParam           | *       | [https://web.expasy.org/protparam/](https://web.expasy.org/protparam/) |
| **Protein-protein interaction network** | STRING              | 10.5    | [https://string-db.org/](https://string-db.org/) |
| **Structure prediction**       | PS2-V2              | 3.0     | [http://ps2.life.nctu.edu.tw/](http://ps2.life.nctu.edu.tw/) |
| **Performance assessment**     | ROC analysis calculator | *     | [http://www.rad.jhmi.edu/jeng/javara/roc/JROCFITi.html](http://www.rad.jhmi.edu/jeng/javara/roc/JROCFITi.html) |

*Information not available*
Table S2. Scores of conserved domain search for 267 HPs of *C. jejuni* strain 81-176 using CDD-Blast, Pfam, HmmScan, SMART and Scanprosite tools.

| No. | Protein ID     | CDD Blast | Pfam | HmmScan | SMART | Scanprosite | Percentage (%) |
|-----|----------------|-----------|------|---------|-------|-------------|----------------|
| 1   | WP_002855595.1 | 1         | 0    | 0       | 0     | 0           | 20             |
| 2   | WP_002868969.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 3   | WP_002866317.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 4   | WP_002824650.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 5   | WP_009881605.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 6   | WP_072238758.1 | 0         | 0    | 0       | 0     | 1           | 20             |
| 7   | WP_024088096.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 8   | WP_002857726.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 9   | WP_002851763.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 10  | WP_002857765.1 | 1         | 0    | 0       | 0     | 1           | 40             |
| 11  | WP_002857751.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 12  | WP_002868767.1 | 1         | 1    | 1       | 0     | 0           | 60             |
| 13  | WP_009881781.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 14  | WP_002868941.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 15  | WP_002857300.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 16  | WP_002851715.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 17  | WP_002854663.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 18  | WP_002851904.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 19  | WP_002857297.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 20  | WP_002870694.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 21  | WP_011812694.1 | 0         | 0    | 0       | 0     | 1           | 20             |
| 22  | WP_002869065.1 | 1         | 0    | 0       | 0     | 0           | 20             |
| 23  | WP_002867117.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 24  | WP_002854172.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 25  | WP_002882716.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 26  | WP_002854628.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 27  | WP_079254179.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 28  | WP_002851686.1 | 1         | 0    | 0       | 0     | 0           | 20             |
| 29  | WP_002868751.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 30  | WP_002859434.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 31  | WP_002854718.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 32  | WP_002854524.1 | 1         | 1    | 1       | 1     | 0           | 80             |
| 33  | WP_002868919.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 34  | WP_002857293.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 35  | WP_002857540.1 | 0         | 0    | 0       | 0     | 0           | 0              |
| 36  | WP_002854351.1 | 1         | 0    | 0       | 0     | 1           | 40             |
|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 117 | WP_002869368.1 | 1 | 1 | 1 | 1 | 0 | 80 |
| 118 | WP_002853267.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 119 | WP_002868960.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 120 | WP_009882583.1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 121 | WP_002853389.1 | 1 | 1 | 1 | 1 | 0 | 80 |
| 122 | WP_002854139.1 | 1 | 0 | 0 | 0 | 0 | 20 |
| 123 | WP_002855981.1 | 0 | 0 | 0 | 0 | 1 | 20 |
| 124 | WP_009882608.1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 125 | WP_011812755.1 | 0 | 1 | 1 | 0 | 0 | 40 |
| 126 | WP_009882621.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 127 | WP_002865636.1 | 1 | 1 | 1 | 1 | 0 | 80 |
| 128 | WP_024088204.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 129 | WP_002869126.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 130 | WP_002869124.1 | 1 | 0 | 0 | 0 | 1 | 40 |
| 131 | WP_002853000.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 132 | WP_002853832.1 | 1 | 0 | 0 | 0 | 0 | 20 |
| 133 | WP_002869121.1 | 1 | 0 | 0 | 0 | 0 | 20 |
| 134 | WP_002856015.1 | 1 | 0 | 0 | 0 | 0 | 20 |
| 135 | WP_002855841.1 | 1 | 0 | 0 | 0 | 0 | 20 |
| 136 | WP_002866103.1 | 0 | 0 | 0 | 0 | 1 | 20 |
| 137 | WP_002869049.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 138 | WP_002856019.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 139 | WP_002852822.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 140 | WP_079254190.1 | 1 | 1 | 1 | 1 | 0 | 80 |
| 141 | WP_002856180.1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 142 | WP_002831611.1 | 1 | 1 | 1 | 1 | 0 | 80 |
| 143 | WP_002825005.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 144 | WP_002852900.1 | 1 | 0 | 0 | 0 | 0 | 20 |
| 145 | WP_002790076.1 | 1 | 1 | 1 | 1 | 1 | 100 |
| 146 | WP_002868861.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 147 | WP_002853792.1 | 1 | 1 | 1 | 1 | 0 | 80 |
| 148 | WP_002859287.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 149 | WP_002868857.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 150 | WP_002853180.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 151 | WP_002869072.1 | 1 | 1 | 1 | 1 | 0 | 80 |
| 152 | WP_002866237.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 153 | WP_002869074.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 154 | WP_002869076.1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 155 | WP_002869078.1 | 1 | 0 | 0 | 0 | 0 | 20 |
| 156 | WP_002869097.1 | 1 | 1 | 1 | 1 | 0 | 80 |
|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 197 | WP_002860117.1 | 1 | 1 | 1 | 1 | 0 |
| 198 | WP_079254198.1 | 0 | 0 | 0 | 0 | 0 |
| 199 | WP_002882243.1 | 1 | 0 | 0 | 0 | 1 |
| 200 | WP_002869298.1 | 0 | 0 | 0 | 0 | 0 |
| 201 | WP_002853105.1 | 0 | 0 | 0 | 0 | 0 |
| 202 | WP_002790440.1 | 1 | 0 | 0 | 0 | 0 |
| 203 | WP_002790442.1 | 0 | 0 | 0 | 0 | 0 |
| 204 | WP_002790713.1 | 0 | 0 | 0 | 0 | 0 |
| 205 | WP_002779702.1 | 0 | 0 | 0 | 0 | 0 |
| 206 | WP_002779703.1 | 0 | 0 | 0 | 0 | 0 |
| 207 | WP_002779704.1 | 0 | 1 | 1 | 1 | 0 |
| 208 | WP_002790730.1 | 1 | 0 | 0 | 0 | 0 |
| 209 | WP_002804244.1 | 0 | 1 | 1 | 0 | 0 |
| 210 | WP_011271766.1 | 0 | 0 | 0 | 0 | 0 |
| 211 | WP_002779777.1 | 0 | 0 | 0 | 0 | 0 |
| 212 | WP_002809140.1 | 0 | 0 | 0 | 0 | 0 |
| 213 | WP_002826068.1 | 0 | 0 | 0 | 0 | 0 |
| 214 | WP_002804272.1 | 0 | 0 | 0 | 0 | 0 |
| 215 | WP_011187233.1 | 1 | 1 | 1 | 1 | 1 |
| 216 | WP_011187234.1 | 1 | 0 | 0 | 0 | 1 |
| 217 | WP_002844160.1 | 0 | 0 | 0 | 0 | 0 |
| 218 | WP_011815226.1 | 0 | 0 | 0 | 0 | 0 |
| 219 | WP_011187235.1 | 1 | 1 | 1 | 1 | 0 |
| 220 | WP_002809051.1 | 1 | 0 | 0 | 0 | 0 |
| 221 | WP_002809052.1 | 0 | 0 | 0 | 0 | 0 |
| 222 | WP_002842869.1 | 0 | 0 | 0 | 0 | 0 |
| 223 | WP_011187239.1 | 0 | 0 | 0 | 0 | 1 |
| 224 | WP_032592775.1 | 0 | 0 | 0 | 0 | 20 |
| 225 | WP_002809111.1 | 1 | 1 | 1 | 1 | 0 |
| 226 | WP_002809110.1 | 0 | 0 | 0 | 0 | 0 |
| 227 | WP_002809107.1 | 0 | 0 | 0 | 0 | 0 |
| 228 | WP_002834241.1 | 0 | 0 | 0 | 0 | 0 |
| 229 | WP_002779751.1 | 0 | 0 | 0 | 0 | 0 |
| 230 | WP_002909884.1 | 1 | 0 | 0 | 0 | 0 |
| 231 | WP_008976813.1 | 0 | 0 | 0 | 0 | 0 |
| 232 | WP_002801797.1 | 0 | 0 | 0 | 0 | 0 |
| 233 | WP_011117548.1 | 1 | 1 | 1 | 1 | 0 |
| 234 | WP_011117549.1 | 1 | 1 | 1 | 1 | 0 |
| 235 | WP_011117551.1 | 0 | 0 | 0 | 0 | 0 |
| 236 | WP_002815556.1 | 0 | 0 | 0 | 0 | 0 |
|   |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|
|  237 | WP_011117559.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  238 | WP_011117563.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  239 | WP_010398003.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  240 | WP_024088118.1 | 1   | 0   | 0   | 0   | 0   | 20  |
|  241 | WP_024088119.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  242 | WP_011117567.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  243 | WP_011799391.1 | 0   | 0   | 0   | 0   | 1   | 20  |
|  244 | WP_011117569.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  245 | WP_011117570.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  246 | WP_011117573.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  247 | WP_011117574.1 | 1   | 0   | 0   | 0   | 0   | 20  |
|  248 | WP_011117575.1 | 1   | 1   | 1   | 0   | 0   | 60  |
|  249 | WP_011799393.1 | 1   | 1   | 1   | 1   | 0   | 80  |
|  250 | WP_011117576.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  251 | WP_011117578.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  252 | WP_011117579.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  253 | WP_011117580.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  254 | WP_011799395.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  255 | WP_011117582.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  256 | WP_011117583.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  257 | WP_002815407.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  258 | WP_004306057.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  259 | WP_011117585.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  260 | WP_011117586.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  261 | WP_079254173.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  262 | WP_011117587.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  263 | WP_011117588.1 | 1   | 1   | 1   | 1   | 0   | 80  |
|  264 | WP_011117589.1 | 0   | 0   | 0   | 0   | 1   | 20  |
|  265 | WP_011799397.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  266 | WP_011799398.1 | 0   | 0   | 0   | 0   | 0   | 0   |
|  267 | WP_011117593.1 | 0   | 0   | 0   | 0   | 0   | 0   |

Note: 0 = 0%, 1 = 25%
Table S3. List of annotated functions of 40 proteins with known function from C. jejuni using Pfam, SMART, MOTIF, INTERPROSCAN, CDART, SUPERFAMILY and SVMprot for ROC analysis.

| No. | Protein ID | Protein Name_Known function | PFAM Prediction | PFAM Score | SMART Prediction2 | SMART Score2 | MOTIF Prediction3 | MOTIF Score3 | INTERPROSCAN Prediction4 | INTERPROSCAN Score4 | CDART Prediction5 | CDART Score5 | SUPERFAMILY prediction6 | SUPERFAMILY Score6 | SVMprot prediction7 | SVMprot Score7 |
|-----|------------|------------------------------|----------------|-----------|-------------------|-------------|-------------------|-------------|--------------------------|------------------|----------------|-------------|--------------------------|----------------|----------------|------------------|
| 1   | WP_009881324.1 | DNA polymerase III subunit beta | DNA polymerase III subunit beta | 1 (5) | DNA polymerase III subunit beta | 1 (5) | DNA polymerase III subunit beta | 1 (5) | DNA polymerase III subunit beta | 1 (5) | DNA polymerase III subunit beta | 1 (5) | Zinc-binding, All DNA-binding | 1 (4) |
| 2   | WP_009881354.1 | glutamate synthase | glutamate synthase | 1 (5) | glutamate synthase | 1 (5) | glutamate synthase | 1 (5) | glutamate synthase | 1 (5) | glutamate synthase | 1 (5) | Manganes e-binding, Zinc-binding | 1 (4) |
| 3   | WP_002855601.1 | CTP synthase | CTP synthase | 1 (5) | CTP synthase | 1 (5) | CTP synthase | 1 (5) | CTP synthase | 1 (5) | CTP synthase | 1 (5) | Nitrogenase iron protein-like, Class I glutamine amidotransferases (GAT) | 1 (2) |
| 4   | WP_011812682.1 | cytochrome c biogenesis protein | cytochrome c biogenesis protein | 1 (5) | cytochrome c biogenesis protein | 1 (5) | cytochrome c biogenesis protein | 1 (5) | cytochrome c biogenesis protein | 1 (5) | cytochrome c biogenesis protein | 1 (5) | No result | 0 (2) |
| 5   | WP_009881534.1 | transglycosylase | transglycosylase | 1 (5) | transglycosylase | 1 (5) | transglycosylase | 1 (5) | transglycosylase | 1 (5) | transglycosylase | 1 (5) | All lipid-binding proteins, Transferases - Glycosyltransferases | 1 (2) |
| 6   | WP_009881539.1 | Bcr/CfIA family efflux MFS transporter | Bcr/CfIA family efflux MFS transporter | 1 (5) | Bcr/CfIA family efflux MFS transporter | 1 (5) | Bcr/CfIA family efflux MFS transporter | 1 (5) | Bcr/CfIA family efflux MFS transporter | 1 (5) | Bcr/CfIA family efflux MFS transporter | 1 (5) | Electrochemical Potential-driven transporters - Porters | 1 (5) |
|    |   WP_002 854281.1 | uracil-DNA glycosylase | uracil-DNA glycosylase | 1 (5) | uracil-DNA glycosylase | 1 (5) | uracil-DNA glycosylase | 1 (5) | uracil-DNA glycosylase | 1 (5) | uracil-DNA glycosylase | 1 (5) | uracil-DNA glycosylase | 1 (5) | Transferases - Glycosyltransferases | 1 (5) |
|----|------------------|------------------------|------------------------|-------|------------------------|-------|------------------------|-------|------------------------|-------|------------------------|-------|------------------------|-------|------------------------|-------|
| 8  | WP_002 853939.1  | acetylglutamate kinase | acetylglutamate kinase | 1 (5) | acetylglutamate kinase | 1 (5) | acetylglutamate kinase | 1 (5) | acetylglutamate kinase | 1 (5) | acetylglutamate kinase | 1 (5) | acetylglutamate kinase | 1 (5) | Electrochemical Potential-driven transporters | 1 (4) |
| 9  | WP_002 854336.1  | molybdate ABC transporter permease | molybdate ABC transporter permease | 1 (5) | molybdate ABC transporter permease | 1 (5) | molybdate ABC transporter permease | 1 (5) | molybdate ABC transporter permease | 1 (5) | molybdate ABC transporter permease | 1 (5) | molybdate ABC transporter permease | 1 (5) | Molybdate ABC transporter permease | 1 (5) |
| 10 | WP_002 859852.1  | multidrug efflux SMR transporter | multidrug efflux SMR transporter | 1 (5) | multidrug efflux SMR transporter | 1 (5) | multidrug efflux SMR transporter | 1 (5) | multidrug efflux SMR transporter | 1 (5) | multidrug efflux SMR transporter | 1 (4) | multidrug efflux SMR transporter | 1 (5) | Metal-binding | 1 (3) |
| 11 | WP_009 882032.1  | prephenate dehydratase | prephenate dehydratase | 1 (5) | prephenate dehydratase | 1 (5) | prephenate dehydratase | 1 (5) | prephenate dehydratase | 1 (5) | prephenate dehydratase | 1 (5) | prephenate dehydratase | 1 (5) | Phosphate binding protein-like | 1 (2) |
|    |                  |                        |                        |       |                        |       |                        |       |                        |       |                        |       |                        |       | Zinc-binding, Forming Carbon-Oxygen Bonds | 1 (2) |
| 12 | WP_002 858694.1  | lysine- tRNA ligase | tRNA synthetases | 1 (5) | rRNA synthetases | 1 (5) | rRNA synthetases | 1 (5) | rRNA synthetases | 1 (5) | rRNA synthetases | 1 (5) | rRNA synthetases | 1 (5) | Forming Carbon-Oxygen Bonds | 1 (3) |
| 13 | WP_009 882169.1  | YigZ family protein | UPF0029 | 1 (2) | UPF0029 | 1 (2) | UPF0029 | 1 (2) | UPF0029 | 1 (2) | YigZ family protein | 1 (2) | YigZ family protein | 1 (5) | Transferring Phosphorus-Containing Groups | 1 (2) |
| 14 | WP_002 857383.1  | FAD-binding protein | FAD-binding protein | 1 (5) | FAD-binding protein | 1 (5) | FAD-binding protein | 1 (5) | FAD-binding protein | 1 (5) | Fumarate reductase flavoprotein C-term | 1 (4) | FAD-binding protein | 1 (5) | Acting on the CH-CH group of donors | 1 (4) |
| 15 | WP_002 782934.1  | ribosomal protein S12 | ribosomal protein S12 | 1 (5) | ribosomal protein S12 | 1 (5) | ribosomal protein S12 | 1 (5) | ribosomal protein S12 | 1 (5) | ribosomal protein S12 | 1 (5) | ribosomal protein S12 | 1 (5) | Nucleic acid-binding proteins | 0 (2) |
|    |                  |                        |                        |       |                        |       |                        |       |                        |       |                        |       |                        |       | rRNA-binding proteins | 1 (4) |
| 16 | WP_002 857290.1  | HIT domain-containing protein | HIT domain-containing protein | 1 (5) | HIT domain-containing protein | 1 (5) | HIT domain-containing protein | 1 (5) | HIT domain-containing protein | 1 (5) | HIT domain-containing protein | 1 (5) | HIT domain-containing protein | 1 (5) | Transferases - Acyltransferases | 1 (3) |
|   | ID       | Description                          | Action       | X-Coord | Y-Coord | X-width | Y-width | X-max | Y-max |
|---|----------|--------------------------------------|--------------|---------|---------|---------|---------|-------|-------|
| 17| WP_002 869243.1 | ferrochelatase                      | Proton       | 1       | 1       | 1       | 1       | 1     | 1     |
|   |          | ferrochelatase                      |              | (3)     | 0       | (3)     | 0       |       |       |
| 18| WP_002 854879.1 | flagellar basal body protein FlgB  | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | flagellar basal body protein FlgB  |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 19| WP_002 880964.1 | endolytic transglycosylase MltG     | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | endolytic transglycosylase MltG     |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 20| WP_002 856958.1 | ATP-binding cassette domain-containing protein | No result | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | ATP-binding cassette domain-containing protein |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 21| WP_002 868904.1 | TolC family protein                 | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | TolC family protein                 |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 22| WP_002 869361.1 | carbamoyltransferase HypF           | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | carbamoyltransferase HypF           |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 23| WP_002 869360.1 | hydrogenase formation protein HypD | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | hydrogenase formation protein HypD |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 24| WP_002 869354.1 | aspartate--tRNA ligase              | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | aspartate--tRNA ligase              |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 25| WP_002 869349.1 | MFS transporter                     | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | MFS transporter                     |              | (2)     | (2)     | (2)     | (2)     |       |       |

**Phosphorus-Oxygen Lyases**

|   | ID       | Description                          | Action       | X-Coord | Y-Coord | X-width | Y-width | X-max | Y-max |
|---|----------|--------------------------------------|--------------|---------|---------|---------|---------|-------|-------|
| 17| WP_002 869243.1 | ferrochelatase                      | Proton       | 1       | 1       | 1       | 1       | 1     | 1     |
|   |          | ferrochelatase                      |              | (3)     | 0       | (3)     | 0       |       |       |
| 18| WP_002 854879.1 | flagellar basal body protein FlgB  | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | flagellar basal body protein FlgB  |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 19| WP_002 880964.1 | endolytic transglycosylase MltG     | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | endolytic transglycosylase MltG     |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 20| WP_002 856958.1 | ATP-binding cassette domain-containing protein | No result | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | ATP-binding cassette domain-containing protein |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 21| WP_002 868904.1 | TolC family protein                 | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | TolC family protein                 |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 22| WP_002 869361.1 | carbamoyltransferase HypF           | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | carbamoyltransferase HypF           |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 23| WP_002 869360.1 | hydrogenase formation protein HypD | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | hydrogenase formation protein HypD |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 24| WP_002 869354.1 | aspartate--tRNA ligase              | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | aspartate--tRNA ligase              |              | (2)     | (2)     | (2)     | (2)     |       |       |
| 25| WP_002 869349.1 | MFS transporter                     | No result    | 0       | 0       | 0       | 0       | 0     | 0     |
|   |          | MFS transporter                     |              | (2)     | (2)     | (2)     | (2)     |       |       |
| Gene Accession | Description                               | Entry Type   | Domain   | Superfamily | Patch Type | Mapping Function | Additional Comments |
|----------------|-------------------------------------------|--------------|----------|-------------|-------------|------------------|---------------------|
| WP_002 869372.1 | molecular chaperone DnaK                  | molecular    | 1        | (5)         |             |                  |                     |
| WP_009 882420.1 | nucleotide exchange factor GrpE           | nucleotide   | 1        | (5)         |             |                  |                     |
| WP_002 857174.1 | serine O-acetyltransferase                | Hexapeptide  | 1        | (2)         |             |                  |                     |
| WP_011 812734.1 | ATP-dependent helicase                    | ATP-dependent| 1        | (5)         |             |                  |                     |
| WP_002 867950.1 | alpha/beta hydrolase                     | alpha/beta   | 1        | (5)         |             |                  |                     |
| WP_002 869103.1 | c-type cytochrome                        | c-type cytochrome | 1 | (5) |             |                  |                     |
| WP_011 812744.1 | DNA translocase FtsK                      | DNA translocase FtsK | 1 | (5) |             |                  |                     |
| WP_002 853404.1 | GNAT family N-acetyltransferase          | GNAT family N-acetyltransferase | 1 | (5) |             |                  |                     |
| WP_002 853451.1 | RNA polymerase sigma factor               | RNA polymerase sigma factor | 1 | (5) |             |                  |                     |
|   | RpoD          | RpoD          | RpoD          | RpoD          | RpoD          | RpoD          | RpoD          |   |
|---|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---|
| 35| WP_002856550.1 | potassium transporter TrkA | potassium transporter TrkA | 1 (5) | potassium transporter TrkA | 1 (5) | potassium transporter TrkA | 1 (5) | potassium transporter TrkA | 1 (5) | Glycosyltransferases, Acting on Ester Bonds | 1 (3) |
| 36| WP_002852861.1 | SsrA-binding protein SmpB | SsrA-binding protein SmpB | 1 (5) | SsrA-binding protein SmpB | 1 (5) | SsrA-binding protein SmpB | 1 (5) | SsrA-binding protein SmpB | 1 (5) | RNA-binding proteins | 1 (3) |
| 37| WP_002855885.1 | FAD-binding protein | FAD-binding protein | 1 (5) | FAD-binding protein | 1 (5) | FAD-binding protein | 1 (5) | FAD-binding protein | 1 (5) | Acting on the CH-OH group of donors, Manganese-binding | 1 (3) |
| 38| WP_002856003.1 | riboflavin synthase | riboflavin synthase | 1 (5) | riboflavin synthase | 1 (5) | riboflavin synthase | 1 (5) | riboflavin synthase | 1 (5) | Transferring Alkyl or Aryl Groups, Other than Methyl Groups | 1 (3) |
| 39| WP_002855731.1 | bacterioheme merythrin | bacterioheme merythrin | 1 (5) | bacterioheme merythrin | 1 (5) | bacterioheme merythrin | 1 (5) | bacterioheme merythrin | 1 (5) | Zinc-binding, All DNA-binding domains | 1 (3) |
| 40| WP_002869409.1 | GDP-L-fucose synthase | NAD-dependent epimerase | 0 (2) | GDP-L-fucose synthase | 1 (5) | GDP-L-fucose synthase | 1 (5) | Short-chain dehydrogenases/reductases (SDR) | 0 (2) | ADP-binding Rossmann-fold domains | 0 (2) | server error | 0 (2) |
| No. | Protein ID          | PFAM                        | SMART          | MOTIF                          | INTERPROSCAN                        | CDART                      | SUPERFAMILY | SVMProt          |
|-----|---------------------|-----------------------------|----------------|--------------------------------|--------------------------------------|---------------------------|--------------|------------------|
| 1   | WP_002868 767.1     | TolB amino-terminal domain  | CsgG           | Curli production assembly, transport component CsgG | Curli production assembly, transport component CsgG | TolB amino-terminal domain | CC0632-like | All lipid-binding proteins |
| 2   | WP_002854 524.1     | Chemotaxis phosphatase CheX | CheX           | Chemotaxis phosphatase CheX   | Chemotaxis phosphatase cheX-like domain | CheC-like family           | CheC-like    | Transferring Phosphorus-Containing Groups |
| 3   | WP_009882 162.1     | SprA-related family         | SprA-related   | SprA-related family           | SprA-related family                  | SprA-related family       | No result    | Zinc-binding      |
| 4   | WP_010790 856.1     | Pyridoxamine 5'-phosphate oxidase | Pyridox_oxi dase | Pyridoxamine 5'-phosphate oxidase | Pyridoxamine 5'-phosphate oxidase     | Pyridoxine 5'-phosphate (PNP) oxidase-like | PNP-oxidase like | Zinc-binding |
| 5   | WP_009882 239.1     | haemagglutination activity domain | Haemagg_act   | haemagglutination activity domain | Filamentous haemagglutinin, N-terminal | haemagglutination activity domain | Filamentous hemagglutinin, FhaB, secretion domain | All lipid-binding proteins |
| 6   | WP_002854 991.1     | FxsA cytoplasmic membrane protein, FxsA | FxsA          | FxsA cytoplasmic membrane protein | FxsA cytoplasmic membrane protein | No result | No result | TC1.E Channels |
| 7   | WP_002855 029.1     | DNA replication regulator, HobA | HobA           | DNA replication regulator | DNA replication regulator, HobA | DNA replication regulator, HobA | SM11 | All lipid-binding proteins |
| 8   | WP_002868 905.1     | GDSL-like Lipase            | Lipase_GDSL    | GDSL-like Lipase               | GDSL lipase                          | SGHN-hydrolase Superfamily/esterases and lipases | SGHN hydrolase/esterases and lipases | Zinc-binding |
| 9   | WP_002869 356.1     | Divergent polysaccharide deacetylase | Polysacc_de ac_2 | Divergent polysaccharide deacetylase | Divergent polysaccharide deacetylase | Divergent polysaccharide deacetylase | Divergent polysaccharide deacetylase | EC3.1 Hydrolases - Acting on Ester Bonds |
| 10  | WP_002856 929.1     | C4-type zinc ribbon domain  | zf-RING_7      | C4-type zinc ribbon domain     | C4-type zinc ribbon domain           | C4-type zinc ribbon domain | Tropomyosin | All DNA-binding   |
| 11  | WP_002869 028.1     | Esterase-like activity of phytase | Phytase-like   | Esterase-like activity of phytase | Phytase-like domain                  | SdiA-regulated Superfamily | No result | Zinc-binding     |
| 12  | WP_011812 736.1     | DUF234                      | DUF234         | Domain of unknown function DUF234 | DUF4143 Superfamily                  | Restriction endonuclease-like | Zinc-binding |
|   | Accession | Description                                                                 | Accession | Description                                                                 | Accession | Description                                                                 | Accession | Description                                                                 | Accession | Description                                                                 | Accession | Description                                                                 | Accession | Description                                                                 |
|---|-----------|------------------------------------------------------------------------------|-----------|------------------------------------------------------------------------------|-----------|------------------------------------------------------------------------------|-----------|------------------------------------------------------------------------------|-----------|------------------------------------------------------------------------------|-----------|------------------------------------------------------------------------------|-----------|------------------------------------------------------------------------------|
| 13 | WP_002868 809.1 | Ankyrin repeats, Ank_2                                                       | WP_002868 Ankyrin repeats, Ank_2 | Ankyrin repeats- containing domain                                           | ANK Superfamily | Ankyrin repeat                                                             | EC3.2 | Hydrolases - Glycosylases                                                   |
| 14 | WP_002869 368.1 | Type-1V conjugative transfer system mating-pair stabilisation, TraN          | WP_002869 Type-1V conjugative transfer system mating-pair stabilisation, TraN | Type-F conjugative transfer system mating-pair stabilisation protein          | TraN     | Type-1V conjugative transfer system mating-pair stabilisation, TraN         | TB module | All lipid-binding proteins                                                   |
| 15 | WP_009882 583.1 | NLPC_P60                                                                    | WP_009882 NLPC_P60 stabilising domain, N term | NLPC_P60, N-terminal domain                                                   | NlpC/P60 family | NlpC/P60 family                                                          | Forming | Carbon-Oxygen Bonds                                                          |
| 16 | WP_002853 389.1 | Jag_N-termminus                                                              | WP_002853 Jag N-termminus          | Jag N-terminal domain superfamily                                            | Jag N-termminus | No result                                                                  | Transferring | One-Carbon Groups                                                            |
| 17 | WP_009882 608.1 | Adhesin from Campylobacter                                                   | WP_009882 Adhesin from Campylobacter | Adhesin JlpA, Campylobacter                                                  | JLPASA     | JLPASA Superfamily, Adhesin from Campylobacter                               | No result | Zinc-binding                                                               |
| 18 | WP_002856 369.1 | Putative beta-lactamase-inhibitor-like                                       | WP_002856 PepSY-like Putative beta-lactamase-inhibitor-like                     | Putative beta-lactamase-inhibitor-like, PepSY-like                           | Putative beta-lactamase-inhibitor-like, PepSY-like | BT0923-like, Sodium-binding                                                    |
| 19 | WP_079254 190.1 | Betal-4-N-acetylgalactosaminyltransferase (CgtA)                            | WP_079254 Beta-1,4-N-acetylgalactosaminyltransferase (CgtA) | Beta-1,4-N-acetylgalactosaminyltransferase (CgtA)                           | Beta-1,4-N-acetylgalactosaminyltransferase (CgtA) | No result                                                                  | Magnesium-binding |
| 20 | WP_002856 180.1 | No result                                                                   | WP_002856 No result                | No result                                                                   | No result | Heavy-metal-associated domain                                               | Copper-binding |
| 21 | WP_002831 611.1 | Transcription factor zf-TFIIB                                                 | WP_002831 Transcription factor zfc-finger                                       | Transcription factor zfc-finger                                               | Transcription factor zfc-finger | No result                                                                  | Magnesium-binding |
| 22 | WP_002790 076.1 | Methyl-accepting chemotaxis protein (MCP) signalling domain                 | WP_002790 Methyl-accepting chemotaxis protein (MCP) signalling domain           | Methyl-accepting chemotaxis protein (MCP) signalling domain                  | Methyl-accepting chemotaxis protein (MCP) signalling domain                  | Methyl-accepting chemotaxis protein (MCP) signalling domain                  | P-P bond-hydrosynthesis-driven transporters |
| 23 | WP_002853 792.1 | Plasminogen-binding protein pgPA N-terminal                                  | WP_002853 Plasminogen-binding protein pgPA N-terminal                           | Plasminogen-binding protein pgPA N-terminal                                  | Plasminogen-binding protein pgPA N-terminal                                  | No result                                                                  | Manganese-binding |
| 24 | WP_002869 072.1 | Putative S-adenosyl-L-methionine-dependent methyltransferase                 | WP_002869 Methyltransf _28 Putative S-adenosyl-L-methionine-dependent methyltransferase | S-adenosyl-L-methionine-dependent methyltransferase                          | SAM-dependent methyltransferase, MidA family                             | S-adenosyl-L-methionine-dependent methyltransferases | Zinc-binding |
| 25 | WP_002869 097.1 | MaoC like domain                                                             | WP_002869 MaoC_dehydratases MaoC like domain                                   | MaoC-like dehydratase domain                                                  | Short-chain dehydrogenases                                                 | MaoC-like Transferring Phosphorus- |
| Nr | Accession | Description | Domain | Subunit | Domain | Containing Groups |
|----|-----------|-------------|--------|---------|--------|-------------------|
| 26 | WP_002869 326.1 | Carboxypeptidase controlling helical cell shape catalytic | Peptidase_M99 | Carboxypeptidase controlling helical cell shape catalytic | Metallo-carboxypeptidase, C-terminal domain | Zn-dependent exopeptidases |
| 27 | WP_002869 139.1 | Pyruvate phosphate dikinase, PEP | PPDK_N | Pyruvate phosphate dikinase, PEP | Pyruvate phosphate dikinase, PEP | Forming Carbon-Oxygen Bonds |
| 28 | WP_002869 194.1 | No result | Rod-binding | Rod binding protein | Uncharacterised conserved protein | No result |
| 29 | WP_002869 195.1 | Anti-sigma-28 factor | FlgM | Anti-sigma-28 factor | Anti-sigma-28 factor FlgM superfamily | Sodium-binding |
| 30 | WP_002856 630.1 | PD-(D/E)XK nuclease superfamily | PDDEKX1 | PD-(D/E)XK nuclease superfamily, PDDEKX1 | PD-(D/E)XK endonuclease-like domain, AddAB-type | Forming Carbon-Oxygen Bonds |
| 31 | WP_002855 458.1 | MgtE intracellular N domain | MgtE_N | MgtE intracellular N domain | Flagellar motility protein MotE, a chaperone for MotC folding | ATP-binding cassette (ABC) family |
| 32 | WP_002797 496.1 | Flagellar FliJ protein | FliJ | Flagellar FliJ protein | Flagellar FliJ protein | Magnesium-binding |
| 33 | WP_024088 174.1 | Nitrate reductase delta subunit | Nitrate_red_del | Nitrate reductase delta subunit | Nitrate reductase chaperone | TorD-like |
| 34 | WP_009883 030.1 | AAA domain, putative AbiEII toxin, Type IV TA system, AAA_21 | AAA_21 | AAA domain, putative AbiEII toxin, Type IV TA system, AAA_21 | AAA domain, putative AbiEII toxin, Type IV TA system, AAA_21 | ABC transporter ATPase domain-like |
| 35 | WP_002824 979.1 | putative NADH-ubiquinone oxidoreductase chain E | NADH_UOR_E | putative NADH-ubiquinone oxidoreductase chain E | putative NADH-ubiquinone oxidoreductase chain E | SirA-like |
| 36 | WP_002869 225.1 | DMSO reductase anchor subunit (DmsC) | DmsC | DMSO reductase anchor subunit (DmsC) | DMSO reductase anchor subunit (DmsC) | No result |
| 37 | WP_002856 602.1 | Putative beta-lactamase-inhibitor-like | PepSY_like | Putative beta-lactamase-inhibitor-like | Putative beta-lactamase-inhibitor-like, PepSY-like | BT0923-like |

**Legend:**
- **Domain:** The domain is associated with the protein and indicates its functional role.
- **Subunit:** The subunit is a specific part of the protein.
- **Domain:** The domain is associated with the protein and indicates its functional role.
- **Containing Groups:** The containing groups are specific complexes or pathways associated with the protein.
| No. | Accession | Description | Biochemistry | Function | Remarks  |
|-----|------------|-------------|--------------|----------|---------|
| 38  | WP_002868 888.1 | No result | TPR_2/TPR_8 | Tetrapricleptide repeat, TPR_2 | Lipopolysaccharide biosynthesis regulator YciM, contains six TPR domains and a predicted metal-binding C-terminal domain |
| 39  | WP_002868 880.1 | ABC-type transport auxiliary lipoprotein component | ABC_trans_aux | ABC-type transport auxiliary lipoprotein component | ABC-type transport auxiliary lipoprotein component |
| 40  | WP_009883 121.1 | Flagellar FLIS export co-chaperone | FLIS_cochap | Flagellar FLIS export co-chaperone, HP1076 | Flagellar FLIS export co-chaperone, HP1076 |
| 41  | WP_002860 117.1 | Menaquinone biosynthesis | VitK2biosynth | Menaquinone biosynthesis | Member of the type 2 periplasmic binding fold protein superfamily |
| 42  | WP_002779 704.1 | T-antigen specific domain | Papo_T_anti | T-antigen specific domain | No result |
| 43  | WP_011187 233.1 | Toprim domain | TOPRIM | Toprim domain | Uncharacterized domain associated with phage |
| 44  | WP_011187 235.1 | AAA domain, AAA_25 | AAA_25 | AAA domain, AAA_25 | P-loop containing nucleoside triphosphate hydrolase |
| 45  | WP_002809 111.1 | TrbM | TrbM | TrbM | TrbM Superfamily |
| 46  | WP_011117 548.1 | VirB8 protein | VirB8 | VirB8 protein | Bacterial virulence protein VirB8 |
| 47  | WP_011117 549.1 | Conjugation transfer protein | CagX | Conjugation transfer protein | CagX |
| 48  | WP_011117 575.1 | Type IV secretion system proteins, T4SS | T4SS | Type IV secretion system proteins, T4SS | Type IV secretion system, VirB5 |
| 49  | WP_011799 393.1 | TrbM | TrbM | TrbM | TrbM |
| 50  | WP_011117 588.1 | PemK-like, MazF-like toxin of type II toxin-antitoxin system | PemK_toxin | PemK-like, MazF-like toxin of type II toxin-antitoxin system | mRNA interferase |

Downloaded from https://portlandpress.com/bioscirep/article-pdf/doi/10.1042/BSR20193939/882282/bsr-2019-3939.pdf by guest on 27 May 2020

Bioscience Reports. This is an Accepted Manuscript. You are encouraged to use the Version of Record that will replace this version. The most up-to-date version is available at https://doi.org/10.1042/BSR20193939
S5 Table. Results of the blastp search for similar sequences against non-redundant (nr) database

| No. | Protein ID     | Organism                                    | Query cover | e-value   | Score (bits) | Identity | Product                                           |
|-----|----------------|---------------------------------------------|-------------|-----------|--------------|----------|---------------------------------------------------|
| 1   | WP_002868767.1 | Campylobacter jejuni subsp. jejuni 129-258 | 100%        | 0.0       | 798          | 100%     | hypothetical protein                              |
|     |                | Campylobacter jejuni CVM 41974             | 100%        | 0.0       | 795          | 99%      | hypothetical protein                              |
|     |                | Campylobacter jejuni BJ-CJG85377           | 100%        | 0.0       | 793          | 99%      | hypothetical protein                              |
|     |                | Campylobacter jejuni subsp. jejuni 84-25  | 100%        | 0.0       | 793          | 99%      | hypothetical protein                              |
|     |                | Campylobacter jejuni X                    | 100%        | 0.0       | 792          | 99%      | hypothetical protein                              |
|     | WP_002854524.1 | Campylobacter jejuni subsp. jejuni 327     | 100%        | 5.00E-97  | 285          | 100%     | hypothetical protein                              |
|     |                | Campylobacter jejuni subsp. jejuni CG8486 | 100%        | 6.00E-97  | 284          | 100%     | hypothetical protein                              |
|     |                | Campylobacter jejuni subsp. jejuni M1     | 100%        | 7.00E-97  | 284          | 100%     | hypothetical protein                              |
|     |                | Campylobacter jejuni RM1221               | 100%        | 1.00E-96  | 283          | 100%     | hypothetical protein                              |
|     |                | Campylobacter jejuni subsp. jejuni LMG 23211 | 100%       | 3.00E-96  | 282          | 99%      | hypothetical protein                              |
| 3   | WP_009882162.1 | Campylobacter jejuni subsp. jejuni 81-176-DRH212 | 100%       | 6.00E-174 | 487          | 100%     | hypothetical protein                              |
|     |                | Campylobacter sp. BCW_4319               | 100%        | 1.00E-172 | 484          | 99%      | hypothetical protein                              |
|     |                | Campylobacter sp. BCW_4319               | 100%        | 3.00E-171 | 480          | 97%      | hypothetical protein                              |
|     |                | Campylobacter jejuni subsp. jejuni CF93-6 | 100%        | 3.00E-169 | 475          | 97%      | hypothetical protein                              |
| 4   | WP_010790856.1 | Campylobacter jejuni subsp. jejuni 109     | 100%        | 1.00E-96  | 283          | 99%      | pyridoxamine 5’-phosphate oxidase                 |
|     |                | Campylobacter sp. 109                    | 100%        | 3.00E-96  | 282          | 99%      | pyridoxamine 5’-phosphate oxidase                 |
|     |                | Campylobacter sp. 3                      | 100%        | 5.00E-96  | 281          | 99%      | pyridoxamine 5’-phosphate oxidase                 |
|     |                | Campylobacter sp. BCW_8713              | 100%        | 6.00E-96  | 281          | 99%      | pyridoxamine 5’-phosphate oxidase                 |
|     |                | Campylobacter jejuni subsp. doylei      | 100%        | 6.00E-96  | 281          | 99%      | pyridoxamine 5’-phosphate oxidase                 |
| 5   | WP_009882239.1 | Campylobacter jejuni subsp. jejuni       | 100%        | 0.0        | 1096         | 99%      | filamentous hemagglutinin N-terminal domain-containing protein |
|     |                | Campylobacter jejuni BJ-CJG861114        | 100%        | 0.0        | 1093         | 99%      | filamentous hemagglutinin N-terminal domain-containing protein |
|     |                | Campylobacter jejuni subsp. jejuni 81-176-UMCW9 | 100%       | 0.0        | 1093         | 99%      | filamentous hemagglutinin N-terminal domain-containing protein |
| No. | Accession Number | Organism | Description | Identity | E-Value | Length | Coverage | Function |
|-----|------------------|----------|-------------|----------|---------|--------|----------|----------|
| 6   | WP_002854991.1   | *Campylobacter jejuni* RM1221 | integral membrane protein | 100%     | 8.00E-83 | 247    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 260.94 | integral membrane protein | 100%     | 1.00E-82 | 247    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni H893-13 | integral membrane protein | 100%     | 2.00E-82 | 246    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni D2600 | integral membrane protein | 100%     | 2.00E-82 | 246    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni LMG 23264 | integral membrane protein | 100%     | 2.00E-82 | 246    | 98%      |          |
| 7   | WP_002855029.1   | *Campylobacter jejuni* subsp. jejuni 260.94 | hypothetical protein | 100%     | 2.00E-124 | 357    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni ICDC107001 | hypothetical protein | 100%     | 6.00E-124 | 355    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 129-258 | hypothetical protein | 100%     | 6.00E-124 | 355    | 99%      |          |
|     |                  | *Campylobacter jejuni* K1 | hypothetical protein | 100%     | 6.00E-124 | 355    | 98%      |          |
| 8   | WP_002868905.1   | *Campylobacter jejuni* subsp. jejuni 129-258 | hypothetical protein | 100%     | 0.0 | 787    | 100%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 81-176-DRH212 | hypothetical protein | 100%     | 0.0 | 786    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni CG8486 | hypothetical protein | 100%     | 0.0 | 784    | 99%      |          |
| 9   | WP_002869356.1   | *Campylobacter jejuni* subsp. jejuni 129-258 | polysaccharide deacetylase | 100%     | 0.0 | 718    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 2008-872 | polysaccharide deacetylase | 100%     | 0.0 | 716    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 1997-11 | polysaccharide deacetylase | 100%     | 0.0 | 716    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 1798 | polysaccharide deacetylase | 100%     | 0.0 | 716    | 99%      |          |
|     |                  | *Campylobacter jejuni* K1 | polysaccharide deacetylase | 100%     | 0.0 | 716    | 99%      |          |
| 10  | WP_002856929.1   | *Campylobacter jejuni* RM1221 | zinc ribbon domain protein | 100%     | 3.00E-167 | 470    | 100%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 53 | zinc ribbon domain protein | 100%     | 1.00E-166 | 468    | 99%      |          |
|     |                  | *Campylobacter jejuni* subsp. jejuni 327 | zinc ribbon domain protein | 100%     | 1.00E-166 | 468    | 99%      |          |
|     |                  | *Campylobacter jejuni* CVM 41910 | zinc ribbon domain protein | 100%     | 1.00E-166 | 468    | 99%      |          |
|     |                  | *Campylobacter jejuni* CVM 41927 | zinc ribbon domain protein | 100%     | 2.00E-166 | 468    | 99%      |          |
| 11  | WP_002869028.1   | *Campylobacter jejuni* subsp. jejuni 81-176-DRH212 | glycerophosphodiester | 100%     | 0.0 | 884    | 100%      |          |
| Accession | Description | Identity | Similarity | E-value | Description |
|-----------|-------------|----------|------------|---------|-------------|
| 176-DRH212 | Campylobacter jejuni subsp. jejuni 81-176-UMCW7 | 100% | 0.0 | 883 | 99% | phosphodiesterase |
|           | Campylobacter sp. BCW_4319 | 100% | 0.0 | 882 | 99% | phosphodiesterase |
|           | Campylobacter sp. 1 | 100% | 0.0 | 882 | 99% | phosphodiesterase |
|           | Campylobacter jejuni 30318 | 100% | 0.0 | 881 | 99% | phosphodiesterase |
| 12 WP_011812736.1 | Campylobacter sp. 114 | 100% | 0.0 | 565 | 100% | hypothetical protein |
|           | Campylobacter sp. BCW_8709 | 100% | 0.0 | 565 | 99% | hypothetical protein |
|           | Campylobacter jejuni subsp. jejuni 260.94 | 100% | 0.0 | 564 | 99% | hypothetical protein |
| 13 WP_002868809.1 | Campylobacter jejuni subsp. jejuni 129-258 | 100% | 0.0 | 816 | 100% | ankyrin repeat-containing protein |
|           | Campylobacter jejuni subsp. jejuni 81-176-UMCW7 | 100% | 0.0 | 815 | 99% | ankyrin repeat-containing protein |
|           | Campylobacter jejuni subsp. jejuni 1213 | 100% | 0.0 | 814 | 99% | ankyrin repeat-containing protein |
|           | Campylobacter jejuni K1 | 100% | 0.0 | 813 | 99% | ankyrin repeat-containing protein |
|           | Campylobacter jejuni subsp. jejuni 1997-11 | 100% | 0.0 | 813 | 99% | ankyrin repeat-containing protein |
| 14 WP_002869368.1 | Campylobacter jejuni subsp. jejuni 81-176-DRH212 | 100% | 7.00E-103 | 300 | 100% | mating pair stabilization protein |
|           | Campylobacter sp. BCW_4319 | 100% | 2.00E-101 | 296 | 98% | mating pair stabilization protein |
|           | Campylobacter jejuni subsp. jejuni HB93-13 | 100% | 1.00E-100 | 294 | 98% | mating pair stabilization protein |
|           | Campylobacter jejuni K1 | 100% | 2.00E-100 | 294 | 98% | mating pair stabilization protein |
|           | Campylobacter jejuni CVM 41974 | 100% | 4.00E-100 | 293 | 97% | mating pair stabilization protein |
| 15 WP_009882583.1 | Campylobacter jejuni subsp. jejuni LMG 9217 | 100% | 0.0 | 905 | 99% | SH3_6 and SH3_7 domain-containing protein |
|           | Campylobacter sp. BCW_7460 | 100% | 0.0 | 904 | 99% | SH3_6 and SH3_7 domain-containing protein |
|           | Campylobacter jejuni subsp. jejuni 129-258 | 100% | 0.0 | 904 | 99% | SH3_6 and SH3_7 domain-containing protein |
|   | Accession       | Species         | Id   | Description                                                                 |
|---|-----------------|-----------------|------|----------------------------------------------------------------------------|
| 16| WP_002853389.1  | Campylobacter jejuni subsp. jejuni CF93-6 | 100% | 902 99% SH3_6 and SH3_7 domain-containing protein |
|   |                 |                 | 100% | 0.0 99% SH3_6 and SH3_7 domain-containing protein |
|   |                 | Campylobacter jejuni subsp. jejuni 84-25 | 100% | 99% RNA-binding protein |
|   |                 | Campylobacter jejuni subsp. jejuni 1336 | 100% | 99% RNA-binding protein |
|   |                 | Campylobacter jejuni subsp. jejuni JA3902 | 100% | 99% RNA-binding protein |
|   |                 | Campylobacter jejuni subsp. jejuni 305 | 100% | 99% RNA-binding protein |
| 17| WP_009882608.1  | Campylobacter jejuni subsp. jejuni 81-176-UMCW7 | 100% | 727 99% lipoprotein |
|   |                 | Campylobacter jejuni subsp. jejuni 81-176-UMCW9 | 100% | 727 99% lipoprotein |
|   |                 | Campylobacter sp. BCW_4319 | 100% | 725 99% lipoprotein |
|   |                 | Campylobacter jejuni subsp. jejuni HB93-13 | 100% | 724 99% lipoprotein |
|   |                 | Campylobacter jejuni subsp. jejuni 2008-894 | 100% | 724 99% lipoprotein |
| 18| WP_002856369.1  | Campylobacter jejuni subsp. jejuni 260.94 | 100% | 272 99% Putative beta-lactamase-inhibitor-like, PepSY-like |
|   |                 | Campylobacter jejuni subsp. jejuni 81116 | 100% | 271 99% Putative beta-lactamase-inhibitor-like, PepSY-like |
|   |                 | Campylobacter jejuni subsp. jejuni M1 | 100% | 270 99% Putative beta-lactamase-inhibitor-like, PepSY-like |
|   |                 | Campylobacter jejuni subsp. jejuni ICDCCJ07001 | 100% | 270 99% Putative beta-lactamase-inhibitor-like, PepSY-like |
|   |                 | Campylobacter jejuni subsp. jejuni 327 | 100% | 269 99% Putative beta-lactamase-inhibitor-like, PepSY-like |
| 19| WP_079254190.1  | Campylobacter jejuni subsp. jejuni 81-176 | 100% | 110 100% hypothetical protein |
|   |                 | Campylobacter sp. USS54 | 100% | 110 100% hypothetical protein |
| 20| WP_002856180.1  | Campylobacter jejuni subsp. jejuni 81-176-DRH212 | 100% | 124 100% heavy-metal-associated domain |
|   |                 | Campylobacter jejuni subsp. jejuni 81-176-UMCW7 | 100% | 123 98.44% heavy-metal-associated domain |
|     | Accession Number | Organism Name                      | Identity | Similarity | Length | E-value | Domain Description                      |
|-----|------------------|------------------------------------|----------|------------|--------|---------|----------------------------------------|
| 21  | WP_002831611.1   | Campylobacter jejuni subsp. jejuni LMG 23211 | 100%     | 96.88%     | 122    | 3.00E-35 | heavy-metal-associated domain          |
|     |                  | Campylobacter sp. CH186             | 100%     | 98.85%     | 185    | 3.00E-35 | hypothetical protein                   |
|     |                  | Campylobacter sp. US55              | 100%     | 98.85%     | 184    | 6.00E-35 | hypothetical protein                   |
|     | WP_002790076.1   | Campylobacter jejuni subsp. jejuni LMG 9872 | 100%     | 98.85%     | 185    | 3.00E-35 | hypothetical protein                   |
|     |                  | Campylobacter sp. US53              | 100%     | 98.85%     | 184    | 7.00E-35 | hypothetical protein                   |
| 22  | WP_002790076.1   | Campylobacter jejuni subsp. jejuni LMG 9872 | 100%     | 99.56%     | 918    | 0.0     | methyl-accepting chemotaxis protein    |
|     |                  | Campylobacter jejuni subsp. jejuni 2008-1025 | 100%     | 99.78%     | 918    | 0.0     | methyl-accepting chemotaxis protein    |
|     |                  | Campylobacter jejuni subsp. jejuni 260.94 | 100%     | 99.78%     | 918    | 0.0     | methyl-accepting chemotaxis protein    |
| 23  | WP_00283792.1    | Campylobacter jejuni CVM 41973      | 100%     | 100.00%    | 486    | 1.00E-173 | exporting protein                      |
|     |                  | Campylobacter jejuni CVM 41910      | 100%     | 99.59%     | 486    | 2.00E-173 | exporting protein                      |
|     |                  | Campylobacter jejuni CVM 41922      | 100%     | 99.59%     | 485    | 2.00E-173 | exporting protein                      |
|     |                  | Campylobacter jejuni CVM 41914      | 100%     | 99.59%     | 484    | 3.00E-173 | exporting protein                      |
|     |                  | Campylobacter jejuni CVM 41936      | 100%     | 99.59%     | 484    | 6.00E-173 | exporting protein                      |
| 24  | WP_002869072.1   | Campylobacter sp. BCW_4319          | 100%     | 100.00%    | 629    | 0.0     | hypothetical protein                   |
|     |                  | Campylobacter jejuni subsp. jejuni str. RM3420 | 100%     | 99.68%     | 627    | 0.0     | hypothetical protein                   |
|     |                  | Campylobacter jejuni BJ-CJD101      | 100%     | 99.68%     | 627    | 0.0     | hypothetical protein                   |
|     |                  | Campylobacter jejuni CVM 41974      | 100%     | 99.37%     | 623    | 0.0     | hypothetical protein                   |
|     |                  | Campylobacter jejuni subsp. jejuni 140-16 | 100%     | 98.10%     | 621    | 0.0     | hypothetical protein                   |
| 25  | WP_002869097.1   | Campylobacter jejuni subsp. jejuni LMG 9872 | 100%     | 100.00%    | 895    | 0.0     | MaoC like domain                      |
|     |                  | Campylobacter sp. BCW_4319          | 100%     | 99.78%     | 892    | 0.0     | MaoC like domain                      |
|     |                  | Campylobacter sp. US54              | 100%     | 99.78%     | 892    | 0.0     | MaoC like domain                      |
| Accession     | Description                                      | Identity | Similarity | E-value | Description                                      |
|---------------|--------------------------------------------------|----------|------------|---------|--------------------------------------------------|
| WP_002869326.1| Campylobacter jejuni CVM 41974                   | 100%     | 99.34%     | 0.0     | MaoC like domain                                 |
| WP_002869139.1| Campylobacter jejuni subsp. jejuni 129-258      | 100%     | 100.00%    | 0.0     | hypothetical protein                             |
| WP_002869195.1| Campylobacter jejuni subsp. jejuni 129-258      | 100%     | 100.00%    | 5.00E-37| flagellar biosynthesis anti-sigma factor FlgM   |
| WP_002856630.1| Campylobacter jejuni subsp. jejuni str. RM3420  | 100%     | 100.00%    | 0.0     | helicase AddB                                    |
| WP_002855458.1| Campylobacter jejuni RM1221                     | 100%     | 100.00%    | 6.00E-117| nucleosidase                                     |
|               | Campylobacter jejuni subsp. jejuni S3           | 100%     | 99.42%     | 9.00E-117| nucleosidase                                     |
|               | Campylobacter jejuni CJ2                        | 100%     | 99.42%     | 2.00E-116| nucleosidase                                     |
|               | Campylobacter jejuni 20176                      | 100%     | 99.42%     | 2.00E-116| nucleosidase                                     |
| Accession   | Description                       | Identity | E-value | Length | Coverage | Description                      |
|-------------|-----------------------------------|----------|---------|--------|----------|----------------------------------|
| WP_002797496.1 | Campylobacter jejuni RM1221       | 100%     | 2.00E-116 | 336    | 99.42%   | nucleosidase                     |
| WP_024088174.1 | Campylobacter jejuni subsp. jejuni M129 | 100%     | 2.00E-95  | 280    | 100.00%  | hypothetical protein             |
|              | Campylobacter jejuni RM1221       | 100%     | 2.00E-95  | 279    | 99.30%   | hypothetical protein             |
|              | Campylobacter coli 15-537360      | 100%     | 6.00E-95  | 280    | 100.00%  | hypothetical protein             |
|              | Campylobacter jejuni subsp. jejuni R14 | 100%     | 8.00E-95  | 279    | 99.30%   | hypothetical protein             |
|              | Campylobacter jejuni subsp. jejuni M129 | 100%     | 8.00E-95  | 279    | 99.30%   | hypothetical protein             |
| WP_009883030.1 | Campylobacter jejuni 81-176-UMCW7 | 100%     | 0.0      | 865    | 100.00%  | ATP/GTP-binding protein          |
| WP_002824979.1 | Campylobacter jejuni subsp. doylei | 100%     | 5.00E-45  | 148    | 100.00%  | NADH-ubiquinone oxidoreductase   |
|              | Campylobacter jejuni subsp. jejuni 81116 | 100%     | 9.00E-45  | 147    | 98.67%   | NADH-ubiquinone oxidoreductase   |
|              | Campylobacter jejuni subsp. jejuni S3 | 100%     | 9.00E-45  | 147    | 98.67%   | NADH-ubiquinone oxidoreductase   |
|              | Campylobacter jejuni subsp. jejuni PT14 | 100%     | 9.00E-45  | 147    | 98.67%   | NADH-ubiquinone oxidoreductase   |
|              | Campylobacter coli CVM 41944      | 100%     | 0.0      | 865    | 100.00%  | ATP/GTP-binding protein          |
|              | Peptoniphilus sp. HMSC075B08      | 100%     | 0.0      | 863    | 99.77%   | ATP/GTP-binding protein          |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni 81-176-UMCW7 | 100%     | 0.0      | 865    | 100.00%  | ATP/GTP-binding protein          |
|              | Campylobacter jejuni subsp. jejuni 81-176-UMCW7 | 100%     | 0.0      | 865    | 100.00%  | ATP/GTP-binding protein          |
|              | Campylobacter jejuni subsp. jejuni 81-176-UMCW7 | 100%     | 0.0      | 863    | 99.77%   | ATP/GTP-binding protein          |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
|              | Campylobacter jejuni subsp. jejuni D42a | 100%     | 5.00E-166 | 467    | 99.58%   | formate dehydrogenase-specific chaperone |
| Accession | Description                                      | Identity | E-value | Length | Percent Identity | Function                                           |
|-----------|--------------------------------------------------|----------|---------|--------|------------------|----------------------------------------------------|
| 36 WP_002856602.1 | Campylobacter jejuni subsp. jejuni HB93-13         | 100%     | 0.0     | 574    | 99.65%           | dimethylsulfoxide reductase                         |
|           | Campylobacter jejuni CVM 41974                   | 100%     | 0.0     | 573    | 99.31%           | dimethylsulfoxide reductase                         |
| 36 WP_002856602.1 | Campylobacter jejuni subsp. jejuni 81116         | 100%     | 2.00E-94| 278    | 100%             | periplasmic protein                                |
|           | Campylobacter jejuni subsp. jejuni M1             | 100%     | 3.00E-94| 277    | 99.28%           | periplasmic protein                                |
|           | Campylobacter jejuni subsp. jejuni ICDC107001    | 100%     | 3.00E-94| 277    | 99.28%           | periplasmic protein                                |
|           | Campylobacter jejuni subsp. jejuni str. RM3420    | 100%     | 5.00E-94| 276    | 99.28%           | periplasmic protein                                |
| 37 WP_002868888.1 | Campylobacter jejuni subsp. jejuni 1336          | 100%     | 5.00E-94| 276    | 98.55%           | periplasmic protein                                |
|           | Campylobacter jejuni subsp. jejuni S3             | 100%     | 0.0     | 655    | 100.00%          | periplasmic protein                                |
|           | Campylobacter sp. BCW_6461                       | 100%     | 0.0     | 654    | 99.70%           | periplasmic protein                                |
|           | Campylobacter sp. BCW_4319                       | 100%     | 0.0     | 654    | 99.70%           | periplasmic protein                                |
|           | Campylobacter sp. BCW_6871                       | 100%     | 0.0     | 654    | 99.70%           | periplasmic protein                                |
|           | Campylobacter jejuni subsp. jejuni 1213          | 100%     | 0.0     | 653    | 99.70%           | periplasmic protein                                |
| 38 WP_002868880.1 | Campylobacter jejuni subsp. jejuni 260.94        | 100%     | 3.00E-139| 395    | 100%             | ABC transporter                                    |
|           | Campylobacter jejuni subsp. jejuni 81-176        | 100%     | 1.00E-138| 394    | 98.99%           | ABC transporter                                    |
|           | Campylobacter jejuni K1                          | 100%     | 2.00E-138| 394    | 99.50%           | ABC transporter                                    |
|           | Campylobacter sp. BCW_4319                       | 100%     | 2.00E-138| 394    | 99.50%           | ABC transporter                                    |
|           | Campylobacter sp. US54                           | 100%     | 2.00E-138| 394    | 99.50%           | ABC transporter                                    |
| 39 WP_009883121.1 | Campylobacter jejuni subsp. jejuni 81-176        | 100%     | 4.00E-116| 335    | 100.00%          | hypothetical protein                               |
|           | Campylobacter jejuni 32488                       | 100%     | 9.00E-116| 334    | 99.40%           | hypothetical protein                               |
|           | Campylobacter jejuni subsp. jejuni 260.94        | 100%     | 2.00E-115| 333    | 98.80%           | hypothetical protein                               |
|           | Campylobacter jejuni subsp. jejuni 1854          | 100%     | 2.00E-115| 333    | 98.80%           | hypothetical protein                               |
|           | Campylobacter jejuni subsp. jejuni 1893          | 100%     | 3.00E-115| 332    | 98.80%           | hypothetical protein                               |
| 40 WP_002860117.1 | Campylobacter jejuni CVM 41927                   | 100%     | 0.0     | 583    | 100.00%          | S-ribosylhomocysteine lyase                        |
| Accession     | Species                        | Identity | Expect | Length | Bit Score | BLAST Description               |
|--------------|--------------------------------|----------|--------|--------|-----------|---------------------------------|
| WP_002779704.1 | Campylobacter jejuni subsp. jejuni 81-176 | 100%     | 5.00E-56| 176    | 100.00%   | cpp11 like protein               |
| WP_011187233.1 | Campylobacter jejuni subsp. jejuni 81-176 | 100%     | 0.0    | 828    | 100%      | cpp22 like protein               |
| WP_011187235.1 | Campylobacter jejuni subsp. jejuni 81-176 | 100%     | 0.0    | 826    | 99.75%    | cpp22 like protein               |
| WP_002809111.1 | Campylobacter jejuni subsp. jejuni 81-176 | 100%     | 0.0    | 524    | 100.00%   | conjugal transfer protein TrbM   |
| WP_01117548.1  | Campylobacter jejuni subsp. jejuni IA3902 | 100%     | 3.00E-162| 456    | 100.00%   | virulence protein                |
| Campylobacter jejuni CVM 41900 | Campylobacter jejuni subsp. jejuni 1997-1 | 100%     | 0.0    | 582    | 99.65%    | S-ribosylhomocysteine lyase      |
| Campylobacter sp. CH246 | Campylobacter sp. BCW_6462 | 100%     | 0.0    | 582    | 99.65%    | S-ribosylhomocysteine lyase      |
| Campylobacter sp. BCW_6462 | Campylobacter jejuni subsp. jejuni S3 | 100%     | 8.00E-56| 176    | 100.00%   | cpp22 like protein               |
| Campylobacter coli CVM N29710 | Campylobacter coli 2553 | 98%      | 6.00E-55| 174    | 100.00%   | cpp22 like protein               |
| Campylobacter coli 1148 | Campylobacter jejuni subsp. jejuni M129 | 100%     | 0.0    | 1196   | 100.00%   | cpp26 like protein               |
| Campylobacter coli 317/04 | Campylobacter jejuni subsp. jejuni D42a | 100%     | 0.0    | 1193   | 99.83%    | cpp26 like protein               |
| Campylobacter coli 86119 | Campylobacter jejuni subsp. jejuni IA3902 | 100%     | 9.00E-162| 455    | 99.56%    | virulence protein                |
| Campylobacter coli 1148 | Campylobacter jejuni subsp. jejuni IA3902 | 100%     | 0.0    | 523    | 99.61%    | conjugal transfer protein TrbM   |
| Campylobacter coli 317/04 | Campylobacter jejuni subsp. jejuni IA3902 | 100%     | 0.0    | 523    | 99.61%    | conjugal transfer protein TrbM   |
| Campylobacter coli 1148 | Campylobacter jejuni subsp. jejuni IA3902 | 100%     | 0.0    | 521    | 99.61%    | conjugal transfer protein TrbM   |
| Campylobacter sp. BCW_6462 | Campylobacter jejuni subsp. jejuni IA3902 | 100%     | 3.00E-162| 455    | 99.56%    | virulence protein                |
| Protein Name | Percentage | Start | End | Description |
|--------------|------------|-------|-----|-------------|
| Campylobacter coli 132-6 | 100% | 2.00E-161 | 454 | 99.56% virulence protein |
| Campylobacter jejuni X | 100% | 4.00E-160 | 451 | 98.67% virulence protein |
| Campylobacter coli RM1875 | 100% | 6.00E-159 | 448 | 97.78% virulence protein |
| WP_011117549.1 | 46 | Campylobacter jejuni subsp. jejuni 81-176 | 100% | 0.0 | 727 | 100.00% type IV secretion system protein VirB9 |
| WP_011117575.1 | 47 | Campylobacter jejuni subsp. jejuni IA3902 | 100% | 0.0 | 725 | 99.72% type IV secretion system protein VirB10 |
| WP_011799393.1 | 48 | Campylobacter jejuni subsp. jejuni IA3902 | 100% | 0.0 | 536 | 100.00% TrbM-like protein |
| WP_011117588.1 | 49 | Campylobacter jejuni subsp. jejuni 81-176 | 100% | 7.00E-88 | 261 | 100.00% toxin-antitoxin system protein |

Helicobacter canis NCTC 12740 | 71% | 5.00E-61 | 191 | 100.00% toxin-antitoxin system protein |
### S6 Table. Result of essential protein prediction using Geptop.

| Sl no | Protein ID            | Score  |
|-------|-----------------------|--------|
| 1     | WP_002854524.1        | 1      |
| 2     | WP_002854991.1        | 0.6004 |
| 3     | WP_002855029.1        | 1      |
| 4     | WP_002868905.1        | 0.9991 |
| 5     | WP_002856929.1        | 0.4441 |
| 6     | WP_011812736.1        | 1      |
| 7     | WP_002869368.1        | 0.7991 |
| 8     | WP_009882583.1        | 0.3995 |
| 9     | WP_002853389.1        | 1      |
| 10    | WP_009882608.1        | 0.799  |
| 11    | WP_002856369.1        | 0.4539 |
| 12    | WP_002831611.1        | 0.3632 |
| 13    | WP_002853792.1        | 0.4539 |
| 14    | WP_002869072.1        | 1      |
| 15    | WP_002869097.1        | 0.4539 |
| 16    | WP_002869326.1        | 0.2739 |
| 17    | WP_002869139.1        | 0.3632 |
| 18    | WP_002856630.1        | 0.3632 |
| 19    | WP_002855458.1        | 0.285  |
| 20    | WP_002797496.1        | 0.285  |
| 21    | WP_024088174.1        | 0.4274 |
| 22    | WP_002824979.1        | 0.4291 |
| 23    | WP_002856602.1        | 0.5699 |
| 24    | WP_002868888.1        | 1      |
| 25    | WP_002868880.1        | 0.285  |
| 26    | WP_002860117.1        | 0.5716 |
| 27    | WP_002779704.1        | 0.4274 |
| 28    | WP_011187233.1        | 0.5699 |
| 29    | WP_011187235.1        | 0.5698 |
| 30    | WP_002809111.1        | 1      |
| 31    | WP_011117575.1        | 1      |
| 32    | WP_011799393.1        | 1      |
S7 Table. List of predicted physicochemical parameters, sub-cellular localization for the HPs from *C. jejuni*.

| No | Protein IDs   | No of Amino acid | MW   | PI  | Extinction coefficient | Instability Index | Classification | Alphabetic index | Grand average of Hydropathicity (GRAVY) | Sub-cellular localization | Signal Peptide (Signal P) | Trans membrane helices prediction | HMMT OP | TMH MM | SOSUI |
|----|---------------|------------------|------|-----|------------------------|-------------------|---------------|-----------------|------------------------------------------|--------------------------|----------------------------|----------------------------------|---------|--------|-------|
| 1  | WP_0028687   | 67.1             | 400  | 9.1 | 1                      | 22350             | 26.24         | Stable          | 87.52                                    | OuterMemb rane            | OuterMembrane              | YES                | No      | No     | Membra ne, 1 TM helix |
| 2  | WP_0028545   | 24.1             | 140  | 4.7 | 5                      | 19285             | 29.26         | Stable          | 93.43                                    | Cytoplasmic              | Cytoplasmic               | No                 | No      | No     | Soluble |
| 3  | WP_0098821   | 62.1             | 241  | 4.9 | 5                      | 11920             | 35.93         | Stable          | 43.9          | -1.31                                   | Extracellular             | Unknown                 | No                 | No      | No     | Soluble |
| 4  | WP_0107908   | 56.1             | 136  | 8.8 | 8                      | 9190              | 28.8          | Stable          | 77.5          | -0.201                                  | Cytoplasmic              | Unknown                 | No                 | No      | No     | Soluble |
| 5  | WP_0098822   | 39.1             | 553  | 4.7 | 9                      | 33935             | 19.93         | Stable          | 82.19         | -0.41                                   | Extracellular             | OuterMemb rane           | YES                | No      | No     | Soluble |
| 6  | WP_0028549   | 91.1             | 129  | 5.3 | 1                      | 11460             | 22.6          | Stable          | 123.88        | 0.924                                   | InnerMemb rane            | CytoplasmicMembrane      | No                 | 3 TM Helices | 3 TM Helice s | Membra ne, 3 TM helix |
| 7  | WP_0028550   | 29.1             | 178  | 4.9 | 3                      | 34045             | 53.4          | Unstable        | 101.4         | -0.07                                   | Cytoplasmic              | Unknown                 | No                 | No      | No     | Soluble |
| 8  | WP_0028689   | 05.1             | 392  | 9.6 | 3                      | 52830             | 28.08         | Stable          | 96.33         | -0.377                                  | OuterMemb rane            | Unknown                 | YES                | No      | No     | Membra ne, 1 TM helix |
| 9  | WP_0028693   | 56.1             | 360  | 7.6 | 8                      | 17880             | 31.74         | Stable          | 98.06         | -0.489                                  | OuterMemb rane            | Cytoplasmic              | No                 | 1 TM Helices | 1 TM Helice s | Membra ne, 1 TM helix |
| 10 | WP_0028569   | 29.1             | 238  | 5.6 | 4                      | 19160             | 40.81         | Unstable        | 90.5          | -0.789                                  | Cytoplasmic              | Cytoplasmic              | No                 | No      | No     | Soluble |
| 11 | WP_0028690   | 28.1             | 441  | 6.5 | 3                      | 41385             | 23.72         | Stable          | 70.27         | -0.744                                  | Cytoplasmic              | Cytoplasmic              | No                 | No      | No     | Soluble |
| 12 | WP_0118127   | 36.1             | 292  | 9.3 | 3                      | 29590             | 30.81         | Stable          | 100.82        | -0.395                                  | Cytoplasmic              | Cytoplasmic              | No                 | No      | No     | Soluble |
| 13 | WP_0028688   | 09.1             | 408  | 5.3 | 9                      | 54125             | 24.69         | Stable          | 92.89         | -0.256                                  | OuterMemb rane            | Unknown                 | YES                | No      | No     | Soluble |
| 14 | WP_0028693   | 68.1             | 150  | 6.5 | 8                      | 7615              | 27.75         | Stable          | 64.33         | -0.486                                  | Cytoplasmic              | OuterMemb rane           | No                 | No      | No     | Soluble |
| 15 | WP_0098825   | 448              | 52423.9 | 9.2 | 6                      | 66365             | 36.19         | Stable          | 85.8          | -0.439                                  | OuterMemb rane            | Unknown                 | No                 | No      | No     | Membra ne |

Downloaded from https://portlandpress.com/bioscirep/article-pdf/doi/10.1042/BSR20193939/882282/bsr-2019-3939.pdf by guest on 27 May 2020
|   | Accession   | Start | End   | Score  | Unstable | Cytoplasmic Region | Membrane Helices | Membrane, Soluble |
|---|-------------|-------|-------|--------|----------|-------------------|-----------------|------------------|
| 16| WP_0028533 | 272   | 87    | 9.0    | Unstable | Cytoplasmic       | 1 TM helix      | Soluble          |
| 17| WP_0098826 | 372   | 61    | 4.8    | Unstable | Cytoplasmic       | TM helix        | Soluble          |
| 18| WP_0028563 | 138   | 31    | 5.0    | Stable   | Cytoplasmic       | Unknown         | Soluble          |
| 19| WP_0792541 | 57    | 31    | 10.39  | Stable   | Cytoplasmic       | 1 TM helix      | Soluble          |
| 20| WP_0028651 | 64    | 55    | 6.0    | Stable   | Cytoplasmic       | Unknown         | Soluble          |
| 21| WP_0028316 | 87    | 35    | 4.4    | Stable   | Cytoplasmic       | Unknown         | Soluble          |
| 22| WP_0027900 | 459   | 29    | 5.2    | Unstable | Cytoplasmic       | 2 TM helices    | Membra ne, Soluble |
| 23| WP_0028537 | 241   | 29    | 5.1    | Stable   | Cytoplasmic       | Unknown         | Soluble          |
| 24| WP_0028690 | 316   | 84    | 6.2    | Stable   | Cytoplasmic       | Unknown         | Soluble          |
| 25| WP_0028690 | 452   | 74    | 9.4    | Stable   | Cytoplasmic       | Unknown         | Soluble          |
| 26| WP_0028693 | 464   | 9     | 7.1    | Stable   | Cytoplasmic       | 1 TM helices    | Membra ne, Soluble |
| 27| WP_0028691 | 779   | 39    | 5.7    | Unstable | Cytoplasmic       | 1 TM helix      | Soluble          |
| 28| WP_0028691 | 65    | 29    | 7.9    | Stable   | Cytoplasmic       | Unknown         | Soluble          |
| 29| WP_0028566 | 788   | 24    | 5.5    | Stable   | Cytoplasmic       | 1 TM helices    | Membra ne, Soluble |
| 30| WP_0028554 | 172   | 46    | 5.3    | Stable   | Cytoplasmic       | 1 TM helices    | Membra ne, Soluble |
| 31| WP_0027974 | 142   | 95    | 9.0    | Unstable | Cytoplasmic       | 1 TM helices    | Soluble          |
| 32| WP_0240881 | 237   | 16    | 5.4    | Unstable | Cytoplasmic       | 1 TM helices    | Soluble          |
| 33| WP_0098830 | 439   | 55    | 5.6    | Unstable | Cytoplasmic       | 1 TM helices    | Soluble          |
|   | Accession | ID | M | % | V | % | B | % | Stability | Location | Subcellular Localization | CD | TM Helices | Location | Subcellular Localization |
|---|-----------|----|---|---|---|---|---|---|-----------|----------|------------------------|----|-------------|----------|------------------------|
| 34 | WP_0028249 | 75 | 8773.2 | 6.2 | 7 | 5500 | 23.41 | Stable | 105.07 | -0.167 | Cytoplasmic | Unknown | No | No | No | Soluble |
| 35 | WP_0028692 | 75 | 32725.77 | 7.1 | 30.26 | Stable | 113.75 | 0.844 | InnerMembrane | CytoplasmicMembrane | No | 8 TM Helices | 8 TM Helices | Membrane, 8 TM helix |
| 36 | WP_0028566 | 75 | 15376.93 | 7.8 | 3 | 11460 | 17.73 | Stable | 98.91 | -0.201 | Periplasmic | Unknown | YES | No | No | Soluble |
| 37 | WP_0028688 | 75 | 39113.6 | 6.2 | 3 | 25370 | 26.75 | Stable | 95.43 | -0.159 | Cytoplasmic | Cytoplasmic | No | 2 TM Helices | 1 TM Helices | Membrane, 8 TM helix |
| 38 | WP_0028688 | 75 | 23051.39 | 8.9 | 5 | 24535 | 44.54 | Unstable | 101.01 | -0.158 | Extracellular | Unknown | No | 2 TM Helices | No | Soluble |
| 39 | WP_0098831 | 75 | 18730.48 | 4.7 | 3 | 6085 | 23.14 | Stable | 99.94 | -0.131 | Cytoplasmic | Unknown | No | No | No | Soluble |
| 40 | WP_0028601 | 75 | 32643.66 | 5.1 | 9 | 37360 | 28.98 | Stable | 103.78 | -0.095 | Cytoplasmic | Cytoplasmic | No | No | No | Soluble |
| 41 | WP_0027797 | 75 | 10623.3 | 7.5 | 3 | 17795 | 57.44 | Unstable | 81.93 | -0.601 | Cytoplasmic | Unknown | No | No | No | Soluble |
| 42 | WP_0111872 | 75 | 47059.95 | 9.1 | 4 | 48025 | 28.47 | Stable | 79.39 | -0.746 | Cytoplasmic | Cytoplasmic | No | No | No | Soluble |
| 43 | WP_0111872 | 75 | 69031.68 | 7.9 | 4 | 51020 | 29.19 | Stable | 91.57 | -0.431 | OuterMembrane | Cytoplasmic | No | No | No | Soluble |
| 44 | WP_0028091 | 75 | 29365.92 | 8.8 | 8 | 39475 | 41.3 | Unstable | 79.84 | -0.534 | Periplasmic | Cytoplasmic | YES | No | No | Soluble |
| 45 | WP_0111175 | 75 | 25915.78 | 7.7 | 8 | 18910 | 24.92 | Stable | 93.16 | -0.18 | OuterMembrane | Unknown | No | 2 TM Helices | 1 TM Helices | Membrane, 1 TM helix |
| 46 | WP_0111175 | 75 | 40874.42 | 8.8 | 3 | 32320 | 34.91 | Stable | 80.06 | -0.613 | OuterMembrane | Unknown | YES | No | No | Soluble |
| 47 | WP_0111175 | 75 | 33294.59 | 5.6 | 7 | 19495 | 34.99 | Stable | 69.18 | -0.692 | Periplasmic | Unknown | YES | 1 TM Helices | 1 TM Helices | Soluble |
| 48 | WP_0117993 | 75 | 30029.93 | 8.5 | 4 | 60445 | 30.85 | Stable | 70.15 | -0.637 | OuterMembrane | Unknown | YES | 1 TM Helices | No | Soluble |
| 49 | WP_0111175 | 75 | 15715.47 | 9.8 | 2 | 16960 | 38.64 | Stable | 90.82 | -0.587 | Cytoplasmic | Unknown | No | 1 TM Helices | 1 TM Helices | Soluble |
Table S8. PPI of the predicted proteins from C. jejuni

| SL | Protein ID          | Interacted protein                                               | Score |
|----|---------------------|------------------------------------------------------------------|-------|
| 1  | WP_002868767.1      | Lipoprotein, putative (207 aa)                                   | 0.945 |
| 2  | WP_002854524.1      | Flagellar motor switch protein FlIN (102 aa)                     | 0.87  |
| 3  | WP_009882162.1      | HIT family protein (120 aa)                                      | 0.634 |
| 4  | WP_010790856.1      | Pyridoxine 5'-phosphate synthase                                 | 0.9   |
| 5  | WP_009882239.1      | Putative outer-membrane protein (508 aa)                         | 0.908 |
| 6  | WP_002854999.1      | Porphobilinogen deaminase (EC 2.5.1.61)                          | 0.861 |
| 7  | WP_002855029.1      | DNA polymerase III, delta prime subunit, homolog (199 aa)        | 0.964 |
| 8  | WP_002868905.1      | Uncharacterized protein (336 aa)                                 | 0.973 |
| 9  | WP_002856929.1      | Membrane protein insertase YidC                                  | 0.889 |
| 10 | WP_002869028.1      | Sulfate-binding protein precursor (348 aa)                       | 0.884 |
| 11 | WP_011812736.1      | Sensor protein ZraS (EC-2.7.13.3) (339 aa)                       | 0.859 |
| 12 | WP_002868809.1      | NADP-dependent 3-hydroxy acid dehydrogenase YdfG (EC-1.1.1.-) (249 aa) | 0.888 |
| 13 | WP_002869368.1      | Hermonuclease precursor (EC-3.1.31.1) (175 aa)                   | 0.533 |
| 14 | WP_009882583.1      | ATP-dependent RecD-like DNA helicase (EC-3.6.4.12) (447 aa)      | 0.674 |
| 15 | WP_002853389.1      | Membrane protein insertase YidC                                  | 0.889 |
| 16 | WP_009882608.1      | ABC transporter glutamine-binding protein GlnH precursor (279 aa) | 0.643 |
| 17 | WP_002856369.1      | Rhomboid protease AarA (EC-3.4.21.105) (157 aa)                  | 0.614 |
| 18 | WP_002856180.1      | Cadmium, cobalt and zinc/H(+)-K(+)-antiporter (316 aa)           | 0.875 |
| 19 | WP_002831611.1      | Uncharacterized protein (64 aa)                                  | 0.681 |
| 20 | WP_002853792.1      | Plasminogen-binding protein PgbB (332 aa)                        | 0.904 |
| 21 | WP_002869072.1      | Mune DD-endopeptidase MepM (EC-3.4.24.-) (273 aa)                | 0.859 |
| 22 | WP_002790076.1      | Chemotaxis protein CheA (EC-2.7.13.3) (769 aa)                   | 0.978 |
| 23 | WP_002853792.1      | Plasminogen-binding protein PgbB (332 aa)                        | 0.904 |
| 24 | WP_002853792.1      | Mune DD-endopeptidase MepM (EC-3.4.24.-) (273 aa)                | 0.859 |
| 25 | WP_002856630.1      | ATP-dependent helicase/nuclease subunit A (EC-3.1.-.-) (921 aa)  | 0.991 |
| 26 | WP_002855458.1      | Adenylosuccinate synthetase (EC-6.3.4.4)                         | 0.883 |
| 27 | WP_002856630.1      | ATP-dependent helicase/nuclease subunit A (EC-3.1.-.-) (921 aa)  | 0.991 |
| 28 | WP_002855458.1      | Adenylosuccinate synthetase (EC-6.3.4.4)                         | 0.883 |
| 29 | WP_002856630.1      | Adenylosuccinate synthetase (EC-6.3.4.4)                         | 0.883 |
| 30 | WP_002856630.1      | Adenylosuccinate synthetase (EC-6.3.4.4)                         | 0.883 |
| 31 | WP_002856630.1      | Adenylosuccinate synthetase (EC-6.3.4.4)                         | 0.883 |
| 32 | WP_002856630.1      | Adenylosuccinate synthetase (EC-6.3.4.4)                         | 0.883 |
|   | Accession     | Description                                                                 | Similarity |
|---|---------------|------------------------------------------------------------------------------|------------|
| 38 | WP_002868880.1 | ABC transporter, periplasmic substrate-binding protein, putative (296 aa)   | 0.867      |
| 39 | WP_009883121.1 | Lipoprotein, putative (199 aa)                                              | 0.676      |
| 40 | WP_002860117.1 | Aminodeoxyfutalosine synthase (EC-2.5.1.-)                                 | 0.988      |
| 41 | WP_002779704.1 | No result                                                                   | NA         |
| 42 | WP_011187233.1 | No result                                                                   | NA         |
| 43 | WP_011187235.1 | No result                                                                   | NA         |
| 44 | WP_002809111.1 | No result                                                                   | NA         |
| 45 | WP_011117548.1 | No result                                                                   | NA         |
| 46 | WP_011117549.1 | VirB10                                                                      | 0.994      |
| 47 | WP_011117575.1 | No result                                                                   | NA         |
| 48 | WP_011799393.1 | No result                                                                   | NA         |
| 49 | WP_011117588.1 | No result                                                                   | NA         |
| SL No | Accession No | Templates | Domain and function in (PS)2-v2 | % Identity | Organism | Method | Resolution | R-value free | R-value work |
|-------|--------------|-----------|---------------------------------|------------|----------|--------|------------|-------------|-------------|
| 1     | WP_002868767.1 | 1efcA     | PROTEIN (ELONGATION FACTOR)     | 51.58%     | Escherichia coli | X-RAY DIFFRACTION | 2.05 Å    | 0.268       | 0.203       |
| 2     | WP_002854524.1 | 1squB     | Chemotaxis phosphatase CheX     | 71.52%     | Thermotoga maritima | X-RAY DIFFRACTION | 2.4 Å     | 0.320       | 0.239       |
| 3     | WP_009882162.1 | No results |                                  |            |           |        |            |             |             |
| 4     | WP_010790856.1 | 2ig6A     | NimC/NimA family protein        | 87.88%     | Clostridium acetobutylicum | X-RAY DIFFRACTION | 1.8 Å     | 0.198       | 0.168       |
| 5     | WP_009882239.1 | No results |                                  |            |           |        |            |             |             |
| 6     | WP_002854991.1 | 1oedB     | ACETYLCHOLINE RECEPTOR PROTEIN, ALPHA CHAIN | 89.19%     | Torpedo marmorata | ELECTRON MICROSCOPY | 4 Å       |             |             |
| 7     | WP_002855029.1 | 2uvpC     | DNA replication regulator HOBA  | 68.48%     | Helicobacter pylori | X-RAY DIFFRACTION | 1.7 Å     | 0.214       | 0.180       |
| 8     | WP_002868905.1 | 1yzfA     | lipase/acylhydrolase            | 79.01%     | Enterococcus faecalis | X-RAY DIFFRACTION | 1.9 Å     | 0.239       | 0.184       |
| 9     | WP_002869356.1 | No results |                                  |            |           |        |            |             |             |
| 10    | WP_002856929.1 | No results |                                  |            |           |        |            |             |             |
| 11    | WP_002869028.1 | 1l3wA     | EP-cadherin                     | 62.14%     | Xenopus laevis | X-RAY DIFFRACTION | 3.08 Å    | 0.276       | 0.243       |
| 12    | WP_011812736.1 | 2e52A     | Type II restriction enzyme HindIII | 71.07%     | Haemophilus influenzae | X-RAY DIFFRACTION | 2 Å       | 0.217       | 0.175       |
| 13    | WP_002868809.1 | 1n11A     | Ankyrin                         | 77.69%     | Homo sapiens | X-RAY DIFFRACTION | 2.7 Å     | 0.303       | 0.319       |
| 14    | WP_002869368.1 | 1eemA     | GLUTATHIONE-S-TRANSFERASE       | 67.95%     | Homo sapiens | X-RAY DIFFRACTION | 2 Å       | 0.271       | 0.219       |
| 15    | WP_009882583.1 | No results |                                  |            |           |        |            |             |             |
| 16    | WP_002853389.1 | No results |                                  |            |           |        |            |             |             |
| 17    | WP_009882608.1 | 2cSuA     | RNA LIGASE                      | 54.57%     | Enterobacteria phage T4 | X-RAY DIFFRACTION | 2.21 Å    | 0.258       | 0.198       |
| 18    | WP_002856369.1 | 3dueA     | Putative periplasmic protein    | 88.80%     | Bacteroides vulgatus | X-RAY DIFFRACTION | 1.85 Å    | 0.233       | 0.192       |
| 19    | WP_079254190.1 | No result  |                                  |            |           |        |            |             |             |
| 20    | WP_002856180.1 | 1osdA     | hypothetical protein MerP       | 70.49%     | Cupriavidus metallidurans | X-RAY DIFFRACTION | 2 Å       | 0.268       | 0.192       |
| 21    | WP_002831611.1 | 2a2pA     | Selenoprotein M                 | 64.71%     | Mus musculus | SOLUTION NMR |          |             |             |
| 22    | WP_002790076.1 | 2ch7A     | METHYL-ACCEPTING                | 89.94%     | Thermotoga maritima | X-RAY DIFFRACTION | 2.5 Å     | 0.297       | 0.259       |
| No. | Accession  | Description | Percentage | Species | Technique | Resolution | d-factor | e-factor |
|-----|------------|-------------|------------|---------|-----------|------------|----------|----------|
| 23  | WP_002853792.1 | CHEMOTAXIS PROTEIN | 73.10% | Rhodopseudomonas palustris | X-RAY DIFFRACTION | 2.1 Å | 0.258 | 0.223 |
| 24  | WP_002869072.1 | 1kDa | DUF185 | 73.10% | Rhodopseudomonas palustris | X-RAY DIFFRACTION | 2.1 Å | 0.258 | 0.223 |
| 25  | WP_002869097.1 | No result | | | | | | |
| 26  | WP_002869326.1 | No results | | | | | | |
| 27  | WP_002869139.1 | No results | | | | | | |
| 28  | WP_002869195.1 | No result | | | | | | |
| 29  | WP_002856630.1 | 1w36F | EXODEOXYRIBONUCLEASE V BETA CHAIN | 62.78% | Escherichia coli | X-RAY DIFFRACTION | 3.1 Å | 0.296 | 0.242 |
| 30  | WP_002855458.1 | 1c1gA | TROPOMYOSIN | 86.59% | Sus scrofa | X-RAY DIFFRACTION | 7 Å | 0.316 | 0.237 |
| 31  | WP_002797496.1 | 2efrA | General control protein GCN4 and Tropomyosin 1 alpha chain | 91.67% | Saccharomyces cerevisiae, Oryctolagus cuniculus | X-RAY DIFFRACTION | 1.8 Å | 0.316 | 0.237 |
| 32  | WP_024088174.1 | 1n1cA | TorA specific chaperone | 84.46% | Shewanella massilia | X-RAY DIFFRACTION | 2.4 Å | 0.255 | 0.224 |
| 33  | WP_009883030.1 | No results | | | | | | |
| 34  | WP_002824979.1 | 1r8sA | ADP-ribosylation factor 1 | 77.78% | Bos taurus, Homo sapiens | X-RAY DIFFRACTION | 1.46 Å | 0.170 | 0.159 |
| 35  | WP_002869225.1 | 2dyrA | Cytochrome c oxidase subunit 1 | 80.54% | Bos taurus | X-RAY DIFFRACTION | 1.8 Å | 0.227 | 0.202 |
| 36  | WP_002856602.1 | 3db7A | putative calcium-regulated periplasmic protein | 88.00% | Bacteroides thetaiotaomicron | X-RAY DIFFRACTION | 1.4 Å | 0.198 | 0.160 |
| 37  | WP_002868888.1 | No results | | | | | | |
| 38  | WP_002868880.1 | 2iqiF | Hypothetical protein XCC0632 | 82.94% | Xanthomonas campestris | X-RAY DIFFRACTION | 2.7 Å | 0.275 | 0.209 |
| 39  | WP_009883121.1 | 1quuA | HUMAN SKELETAL MUSCLE ALPHA-ACTININ 2 | 76.51% | Homo sapiens | X-RAY DIFFRACTION | 2.5 Å | 0.310 | 0.229 |
| 40  | WP_002860117.1 | 3ebmA | hypothetical protein AF1704 | 79.92% | Archaeoglobus fulgidus | X-RAY DIFFRACTION | 2.3 Å | 0.265 | 0.212 |
| 41  | WP_002779704.1 | 3cm5A | RABPHILIN-3A | 72.22% | Rattus norvegicus | X-RAY DIFFRACTION | 1.28 Å | 0.194 |
| 42  | WP_011117231.1 | 2au3A | DNA primase | 58.81% | Aquifex aeolicus | X-RAY DIFFRACTION | 2 Å | 0.238 | 0.203 |
| 43  | WP_011117235.1 | No results | | | | | | |
| 44  | WP_002809111.1 | 3ec1A | YqeH GTPase | 53.75% | Geobacillus stearothermophilus | X-RAY DIFFRACTION | 2.36 Å | 0.287 | 0.254 |
| 45  | WP_011117548.1 | No results | | | | | | |
| 46  | WP_011117549.1 | 2ofqA | TraO | 79.38% | Salmonella | SOLUTION NMR | | | |
|   | WP_011117575.1 | 2ch7A | METHYL-ACCEPTING CHEMOTAXIS PROTEIN | 81.99% | Thermostoga maritima | X-RAY DIFFRACTION | 2.5 Å | 0.297 | 0.259 |
|---|----------------|-------|-----------------------------------|--------|----------------------|------------------|-------|-------|-------|
| 47 | WP_011799393.1 | 1zsoB | hypothetical protein              | 73.97% | Plasmodium falciparum | X-RAY DIFFRACTION | 2.17 Å | 0.233 | 0.183 |
| 48 | WP_011117588.1 | 1ne8A | conserved hypothetical protein YDCE | 56.56% | Bacillus subtilis     | X-RAY DIFFRACTION | 2.1 Å  | 0.210 | 0.159 |