Bioscience Reports (2020) **40** BSR20193939
https://doi.org/10.1042/BSR20193939

Research Article

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

Md. Amran Gazi[1], Sultan Mahmud[2], Shah Mohammad Fahim[1], Md. Rezaul Islam[3], Subhasish Das[1], Mustafa Mahfuz[1,4] and Tahmeed Ahmed[1]

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

Correspondence: Md. Amran Gazi (amran.gazi@icddrb.org)

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, subcellular localization, physicochemical characterization, and protein–protein interactions (PPIs). 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 receiver operating characteristic (ROC) curve analysis. Functional and structural predictions and the results from ROC analyses provided the validity of *in-silico* tools used in the present 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.

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 Fisher syndrome [5]. Recent studies also showed an association of *Campylobacter* infections with malnutrition, a condition highly prevalent in developing countries [2].

Received: 20 November 2019
Revised: 17 May 2020
Accepted: 26 May 2020

Accepted Manuscript online: 27 May 2020
Version of Record published: 09 June 2020

Bioscience Reports (2020) **40** BSR20193939
https://doi.org/10.1042/BSR20193939

Although whole genome sequence of *C. jejuni* NCTC has been published, a detailed catalog of prospective virulence is yet to be documented. Its complete genome contains a circular chromosome of 1641481 base pairs with GC content: 30.6%. Several studies since then suggest *C. jejuni* exhibits high genomic diversity across strains. A shotgun DNA microRNA approach revealed 63-kb long unique genomic DNA sequences in another *Campylobacter* strain, *C. jejuni* 81–176 when compared with 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) 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 lacks 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 have 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 play 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 more 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 the present 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 the present study (accessed 27 February 2019).

Functional analysis of HPs

In order to assign the function using the databases depicted in Supplementary Table S1, first 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 help in the categorization of proteins. Analyses of HPs by five webtools revealed the distinct results. To find a composite result, different confidence levels were assigned on the basis of pooled results obtained from five webtools. For instance, if we observed same results from the five distinct tools, the composite score was 100 (percentage of confidence). For downstream analyses, we filtered 50 out of 267 HPs that displayed 60% or above confidence (Supplementary Table S2).

Next, we performed functional assignment of these 50 selected HPs using different tools (Figure 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 $\geq$90% identical, we considered it as homologues to each other. Query cover, score parameters and e-value of every hit are summarized in Supplementary 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].

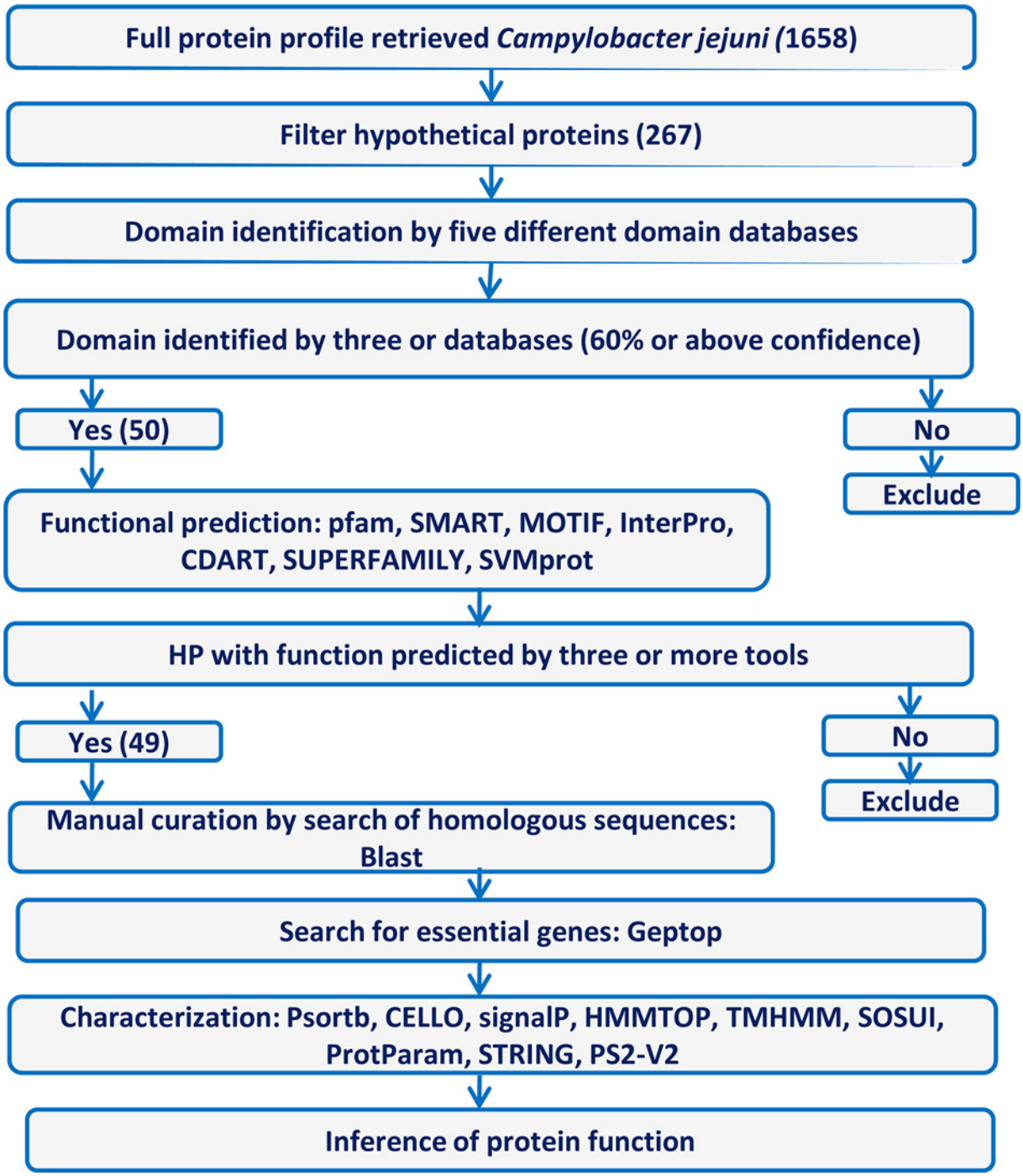

Figure 1. Flow chart showing the overall design of the study

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 subcellular 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

Bioscience Reports (2020) **40** BSR20193939
https://doi.org/10.1042/BSR20193939

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 pairwise 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. First, we selected 40 proteins randomly with known functions of *C. jejuni* (Supplementary Table S3). These proteins were predicted for the functions using the same databases that were 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, the present study focused on the annotation of HPs from *C. jejuni* using assorted but effective bioinformatics tools.

First, 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 have 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. Supplementary 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 (Supplementary 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 (Supplementary Table S5). 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 (Figure 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?2).

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 toward gaining better insight into 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

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	Hemagglutination 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 stabilization, TraN
15	WP_009882583.1	NLPC_P60 stabilizing domain
16	WP_002853389.1	Jag, N-terminal domain superfamily
17	WP_009882608.1	Adhesin from *Campylobacter*
18	WP_002856369.1	Putative β-lactamase-inhibitor-like
19	WP_079254190.1	β-1,4-N-acetylgalactosaminyltransferase (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) signaling 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 β-lactamase-inhibitor-like
37	WP_002868888.1	Tetratricopeptide 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

Bioscience Reports (2020) **40** BSR20193939
https://doi.org/10.1042/BSR20193939

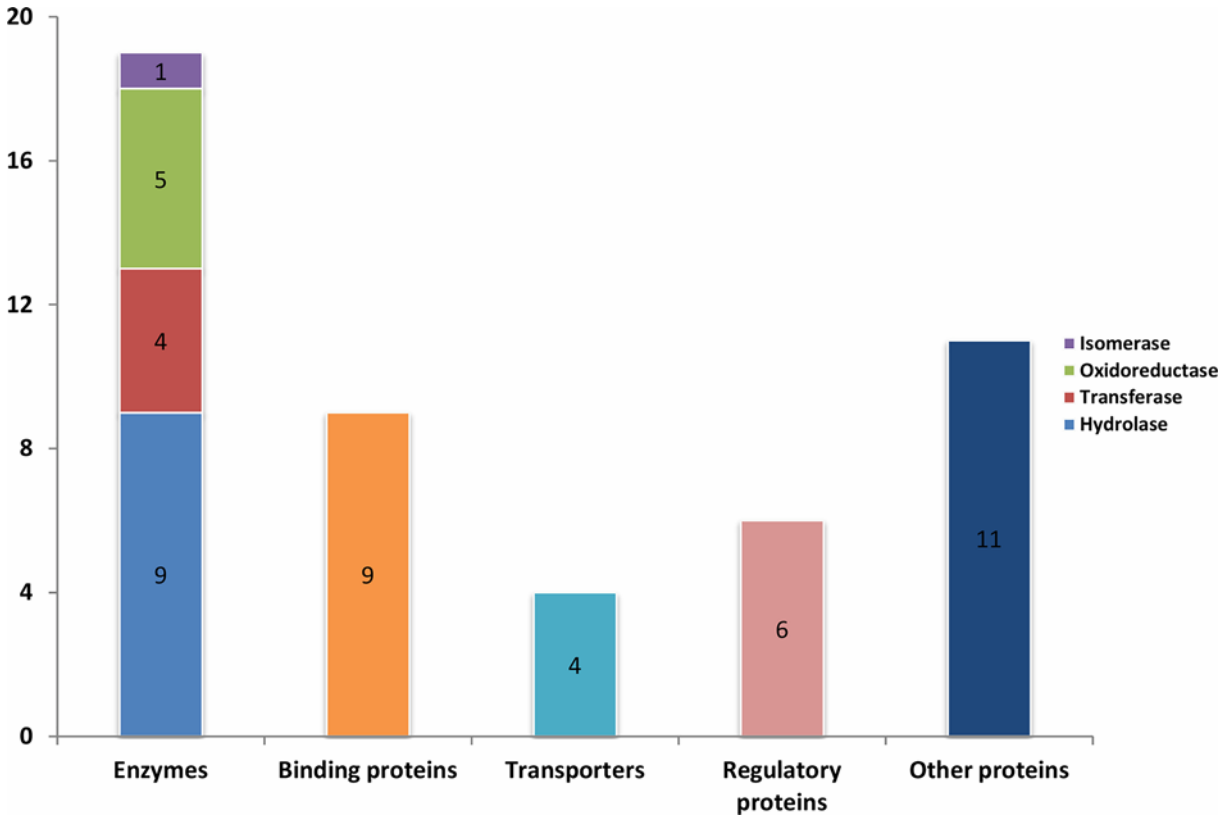

Figure 2. Functional classification of 49 HPs into various groups

Table 2 ROC results of various tools used in the present 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**

for predicting essential genes. In the current study, it was possible to identify 32 essential proteins by using Geptop database (Supplementary 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. DrugBank 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 an 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 sulfoxide 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 represent 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 be 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 a hemagglutination activity domain found in a number of large, repetitive proteins of bacteria. Filamentous hemagglutinin (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 PPI 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 T_4 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 utilize 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 mammals [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].

Bioscience Reports (2020) **40** BSR20193939
https://doi.org/10.1042/BSR20193939

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 transporter E (MgtE), found in eukaryotic proteins. Magnesium (Mg^{2+}) is an essential element for growth and maintenance of living cells where MgtE transports magnesium across the cell membrane [66]. WP_002868880.1 is an ABC-type transport, responsible for outer membrane biosynthesis in bacteria that can be an excellent drug target [67]. WP_002856180.1 is heavy metal-associated (HMA) domain found in a number of detoxification proteins or in heavy metals transport. Proteins that are involved in transporting heavy metals in bacteria, plants, 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 PPIs. 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 applications

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 evolutionarily 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 biocontrol agents to inhibit plant pests. For example, the phenazine pyocyanin contributes to its potential to colonize 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 antimicrobial and anticancer agent [73,74].

Protein WP_024088174.1 is the nitrate reductase that produces 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 which are 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].

Peptidase, 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 hydrolyzes 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 ester compounds of biochemical, food, pharmaceutical, and other biological interests [79].

Other proteins

WP_002856369.1 and WP_002856602.1 was found to be β-lactamase-inhibitor, a group of enzymes responsible for bacterial resistance to β-lactam antibiotics [80]. WP_009883121.1 s ass fla agellar FLiS export co-chaperone. Previously, various FliS-associated proteins in *H. pylori* were identified by a yeast two-hybrid study, but the implications are unknown [81]. Chaperones are usually involved in various important processes such as protein degradation, folding, and polypeptide translocation [81].

At last, 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 from 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 (Supplementary 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 to be biotechnologically important in the present study. Some proteins, essential for pathogenesis of *Campylobacter* have MW ranged from 8773.25 to 39113.6. The pI 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 thermostability [84]. Protein WP_002856369.1, associated with β-lactamase inhibition showed the highest values of 133.14. The GRAVY of protein indicates its hydrophobicity or the interaction with water [85]. In WP_002869028.1, WP_009882583.1, and WP_024088174.1, the scores are among −0.744, −0.439, and −0.393. Moreover, the instability index offers an assumption of the stability of protein *in vitro*. We used cut-off values >40 and <40 to discriminate between stable and unstable proteins, respectively. From our listed proteins, WP_024088174.1 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.

PPI 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 50 unknown proteins for *Streptococcus pneumoniae* [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_024088174.1 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 (Supplementary Table S8).

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 organisms have manifested as surrogates in drug discovery. For example, Nolatrexed, an anticancer 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 the present study. In contrast, nine proteins showed discrepant results and no suitable templates were found for 16 proteins (Supplementary Table S9). Identity of model ranged 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

Bioscience Reports (2020) **40** BSR20193939
https://doi.org/10.1042/BSR20193939

Torpedo marmorata (PDB 1oedB). Both 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 the present 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 the present 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 the present study will be innovative for further *in-vitro/in-vivo* analysis on *C. jejuni*.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

The authors declare that there are no sources of funding to be acknowledged.

Author Contribution

M.A.G. has made substantial contributions to conception, design and drafting the manuscript. S.M., S.M.F., M.R.I. and S.D. participated in the acquisition, analysis and interpretation of data. M.M. and T.A. conceived 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, U.K. (DFID). We gratefully acknowledge these donors for their support and commitment to icddr,b' research efforts.

Abbreviations

ABC , ATP-binding cassette; ANK , ankyrin repeat; AUC, area under the curve; FDA , food and drug administration; GDSL , motif consensus amino acid sequence of Gly, Asp, Ser, and Leu around the active site Ser; GRAVY, grand average of hydropathicity; Hconf, highly confident protein; HMA , heavy-metal-associated; HP, hypothetical protein; MCP , methyl-accepting chemotaxis protein; MgtE, magnesium transporter E; ORF , open reading frame; pdxH, pyridoxamine 5′-phosphate oxidase; pI, isoelectric point; PPI, protein–protein interaction; ROC, receiver operating characteristic; SAM-MT , S-adenosyl-L-methionine-dependent methyltransferase; TPR , tetratricopeptide repeat.

References

1 Kaakoush, N.O., Castaño-Rodríguez, N., Mitchell, H.M. and Man, S.M. (2015) Global epidemiology of Campylobacter infection. *Clin. Microbiol. Rev.* **28**, 687–720, https://doi.org/10.1128/CMR.00006-15

2 Platts-Mills, J.A. and Kosek, M. (2014) Update on the burden of Campylobacter in developing countries. *Curr. Opin. Infect. Dis.* **27**, 444, https://doi.org/10.1097/QCO.0000000000000091
3 Mehla, K. and Ramana, J. (2015) Novel drug targets for food-borne pathogen Campylobacter jejuni: an integrated subtractive genomics and comparative metabolic pathway study. *OMICS* **19**, 393–406, https://doi.org/10.1089/omi.2015.0046
4 Coker, A.O., Isokpehi, R.D., Thomas, B.N., Amisu, K.O. and Obi, C.L. (2002) Human campylobacteriosis in developing countries1. *Emerg. Infect. Dis.* **8**, 237, https://doi.org/10.3201/eid0803.010233
5 Takahashi, M., Koga, M., Yokoyama, K. and Yuki, N. (2005) Epidemiology of Campylobacter jejuni isolated from patients with Guillain-Barré and Fisher syndromes in Japan. *J. Clin. Microbiol.* **43**, 335–339, https://doi.org/10.1128/JCM.43.1.335-339.2005
6 Poly, F., Threadgill, D. and Stintzi, A. (2005) Genomic diversity in Campylobacter jejuni: identification of C. jejuni 81-176-specific genes. *J. Clin. Microbiol.* **43**, 2330–2338, https://doi.org/10.1128/JCM.43.5.2330-2338.2005
7 Parkhill, J., Wren, B., Mungall, K., Ketley, J., Churcher, C., Basham, D. et al. (2000) The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* **403**, 665, https://doi.org/10.1038/35001088
8 Nimrod, G., Schushan, M., Steinberg, D.M. and Ben-Tal, N. (2008) Detection of functionally important regions in "hypothetical proteins" of known structure. *Structure* **16**, 1755–1763, https://doi.org/10.1016/j.str.2008.10.017
9 Shahbaaz, M., ImtaiyazHassan, M. and Ahmad, F. (2013) Functional annotation of conserved hypothetical proteins from Haemophilus influenzae Rd KW20. *PLoS ONE* **8**, e84263, https://doi.org/10.1371/journal.pone.0084263
10 Gazi, M.A., Kibria, M.G., Mahfuz, M., Islam, M.R., Ghosh, P., Afsar, M.N.A. et al. (2016) Functional, structural and epitopic prediction of hypothetical proteins of Mycobacterium tuberculosis H37Rv: an in silico approach for prioritizing the targets. *Gene* **591**, 442–455, https://doi.org/10.1016/j.gene.2016.06.057
11 da Costa, W.L.O., de Aragão Araújo, C.L., Dias, L.M., de Sousa Pereira, L.C., Alves, J.T.C., Araújo, F.A. et al. (2018) 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* **13**, e0198965, https://doi.org/10.1371/journal.pone.0198965
12 Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M. et al. (2016) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368, https://doi.org/10.1093/nar/gkw937
13 Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I. and Stagljar, I. (2015) Fundamentals of protein interaction network mapping. *Mol. Syst. Biol.* **11**, 848, https://doi.org/10.15252/msb.20156351
14 Jez, J.M. (2017) Revisiting protein structure, function, and evolution in the genomic era. *J. Invertebr. Pathol.* **142**, 11–15, https://doi.org/10.1016/j.jip.2016.07.013
15 Gherardini, P.F. and Helmer-Citterich, M. (2008) Structure-based function prediction: approaches and applications. *Brief. Funct. Genom. Proteomics* **7**, 291–302, https://doi.org/10.1093/bfgp/eln030
16 Varma, P.B.S., Adimulam, Y.B. and Kodukula, S. (2015) In silico functional annotation of a hypothetical protein from Staphylococcus aureus. *J. Infect. Public Health* **8**, 526–532, https://doi.org/10.1016/j.jiph.2015.03.007
17 Ravooru, N., Ganji, S., Sathyanarayanan, N. and Nagendra, H.G. (2014) Insilico analysis of hypothetical proteins unveils putative metabolic pathways and essential genes in Leishmania donovani. *Front. Genet.* **5**, 291, https://doi.org/10.3389/fgene.2014.00291
18 Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y. et al. (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226, https://doi.org/10.1093/nar/gku1221
19 Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37, https://doi.org/10.1093/nar/gkr367
20 Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P. and Bork, P. (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234, https://doi.org/10.1093/nar/28.1.231
21 Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R. et al. (2013) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230, https://doi.org/10.1093/nar/gkt1223
22 De Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E. et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365, https://doi.org/10.1093/nar/gkl124
23 Geer, L.Y., Domrachev, M., Lipman, D.J. and Bryant, S.H. (2002) CDART: protein homology by domain architecture. *Genome Res.* **12**, 1619–1623, https://doi.org/10.1101/gr.278202
24 Wilson, D., Madera, M., Vogel, C., Chothia, C. and Gough, J. (2006) The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res.* **35**, D308–D313, https://doi.org/10.1093/nar/gkl910
25 Cai, C., Han, L., Ji, Z.L., Chen, X. and Chen, Y.Z. (2003) SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res.* **31**, 3692–3697, https://doi.org/10.1093/nar/gkg600
26 Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L. et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208, https://doi.org/10.1093/nar/gkp335
27 Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J. et al. (2016) InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199, https://doi.org/10.1093/nar/gkw1107
28 Wei, W., Ning, L.-W., Ye, Y.-N. and Guo, F.-B. (2013) Geptop: a gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. *PLoS ONE* **8**, e72343, https://doi.org/10.1371/journal.pone.0072343
29 Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A. et al. (2010) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39**, D1035–D1041, https://doi.org/10.1093/nar/gkq1126

30 Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, pp. 571–607, Springer, https://doi.org/10.1385/1-59259-890-0:571
31 Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R. et al. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615, https://doi.org/10.1093/bioinformatics/btq249
32 Yu, C.S., Lin, C.J. and Hwang, J.K. (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* **13**, 1402–1406, https://doi.org/10.1110/ps.03479604
33 Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580, https://doi.org/10.1006/jmbi.2000.4315
34 Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379, https://doi.org/10.1093/bioinformatics/14.4.378
35 Tusnady, G.E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850, https://doi.org/10.1093/bioinformatics/17.9.849
36 Armenteros, J.J.A., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S. et al. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol* **37**, 420–423, https://doi.org/10.1038/s41587-019-0036-z
37 Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J. et al. (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452, https://doi.org/10.1093/nar/gku1003
38 Chen, C.-C., Hwang, J.-K. and Yang, J.-M. (2009) 2-v2: template-based protein structure prediction server. *BMC Bioinformatics* **10**, 366, https://doi.org/10.1186/1471-2105-10-366
39 Eng, J. (2017) *ROC Analysis: Web-based Calculator for ROC Curves*, Johns Hopkins University, Baltimore, MD, U.S.A., http://www.jrocfit.org
40 Pearson, W.R. (2013) An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinformatics* **42**, 3.1.8, https://doi.org/10.1002/0471250953.bi0301s42
41 Schnoes, A.M., Brown, S.D., Dodevski, I. and Babbitt, P.C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* **5**, e1000605, https://doi.org/10.1371/journal.pcbi.1000605
42 Gazi, M.A., Mahmud, S., Fahim, S.M., Kibria, M.G., Palit, P., Islam, M.R. et al. (2018) Functional prediction of hypothetical proteins from Shigella flexneri and validation of the predicted models by using ROC curve analysis. *Genomics Inform.* **16**, https://doi.org/10.5808/GI.2018.16.4.e26
43 Zhang, Z. and Ren, Q. (2015) Why are essential genes essential? The essentiality of Saccharomyces genes. *Microb. Cell* **2**, 280, https://doi.org/10.15698/mic2015.08.218
44 Romero, I.G., Ruvinsky, I. and Gilad, Y. (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nat. Rev. Genet.* **13**, 505, https://doi.org/10.1038/nrg3229
45 Hajian-Tilaki, K. (2013) Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J. Int. Med.* **4**, 627
46 Verkhovsky, M.I. and Bogachev, A.V. (2010) Sodium-translocating NADH: quinone oxidoreductase as a redox-driven ion pump. *Biochim. Biophys. Acta* **1797**, 738–746, https://doi.org/10.1016/j.bbabio.2009.12.020
47 Wang, G. and Maier, R.J. (2004) An NADPH quinone reductase of Helicobacter pylori plays an important role in oxidative stress resistance and host colonization. *Infect. Immun.* **72**, 1391–1396, https://doi.org/10.1128/IAI.72.3.1391-1396.2004
48 Kroneck, P.M. and Torres, M.E.S. (2014) The metal-driven biogeochemistry of gaseous compounds in the environment. **14**, 333–335, https://doi.org/10.1007/978-94-017-9269-1
49 Simon, M.I., Borkovich, K.A., Bourret, R.B. and Hess, J.F. (1989) Protein phosphorylation in the bacterial chemotaxis system. *Biochimie* **71**, 1013–1019, https://doi.org/10.1016/0300-9084(89)90105-3
50 Hosfield, D.J., Mol, C.D., Shen, B. and Tainer, J.A. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* **95**, 135–146, https://doi.org/10.1016/S0092-8674(00)81789-4
51 S., K. (2006) Structure, function and mechanisms of action of ATPases from the AAA superfamily of proteins. *Postepy Biochem.* **52**, 330–338
52 Aravind, L., Leipe, D.D. and Koonin, E.V. (1998) Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**, 4205–4213, https://doi.org/10.1093/nar/26.18.4205
53 Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) Introduction to pathogens. *Molecular Biology of the Cell*, 4th, Garland Science
54 Cerveny, L., Straskova, A., Dankova, V., Hartlova, A., Ceckova, M., Staud, F. et al. (2013) Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect. Immun.* **81**, 629–635, https://doi.org/10.1128/IAI.01035-12
55 Jönsson, K., Guo, B.P., Monstein, H.-J., Mekalanos, J.J. and Kronvall, G. (2004) Molecular cloning and characterization of two Helicobacter pylori genes coding for plasminogen-binding proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1852–1857, https://doi.org/10.1073/pnas.0307329101
56 Zhang, J., Zhang, Y., Zhu, L., Suzuki, M. and Inouye, M. (2004) Interference of mRNA function by sequence-specific endoribonuclease PemK. *J. Biol. Chem.* **279**, 20678–20684, https://doi.org/10.1074/jbc.M314284200
57 Makhov, A., Hannah, J., Brennan, M., Trus, B., Kocsis, E., Conway, J. et al. (1994) 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. *J. Mol. Biol.* **241**, 110–124, https://doi.org/10.1006/jmbi.1994.1478
58 Al-Khodor, S., Price, C.T., Kalia, A. and Kwaik, Y.A. (2010) Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol.* **18**, 132–139, https://doi.org/10.1016/j.tim.2009.11.004
59 Kawai, F., Paek, S., Choi, K.-J., Prouty, M., Kanipes, M.I., Guerry, P. et al. (2012) Crystal structure of JlpA, a surface-exposed lipoprotein adhesin of Campylobacter jejuni. *J. Struct. Biol.* **177**, 583–588, https://doi.org/10.1016/j.jsb.2012.01.001

60 Tlapák, H., Rydzewski, K., Schulz, T., Weschka, D., Schunder, E. and Heuner, K. (2017) Functional analysis of the alternative sigma-28 factor FliA and its anti-sigma factor FlgM of the nonflagellated Legionella species L. oakridgensis. *J. Bacteriol.* **199**, e00018–17, https://doi.org/10.1128/JB.00018-17
61 Minamino, T., Chu, R., Yamaguchi, S. and Macnab, R.M. (2000) Role of FliJ in flagellar protein export in Salmonella. *J. Bacteriol.* **182**, 4207–4215, https://doi.org/10.1128/JB.182.15.4207-4215.2000
62 Christie, P.J. and Vogel, J.P. (2000) Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* **8**, 354–360, https://doi.org/10.1016/S0966-842X(00)01792-3
63 Natrajan, G., Noirot-Gros, M.F., Zawilak-Pawlik, A., Kapp, U. and Terradot, L. (2009) The structure of a DnaA/HobA complex from Helicobacter pylori provides insight into regulation of DNA replication in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21115–21120, https://doi.org/10.1073/pnas.0908966106
64 Roujeinikova, A. and Ud-Din AIMS (2017) Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cell. Mol. Life Sci.* **74**, 3293–3303, https://doi.org/10.1007/s00018-017-2514-0
65 Yuan, J., Zweers, J.C., Van Dijl, J.M. and Dalbey, R.E. (2010) Protein transport across and into cell membranes in bacteria and archaea. *Cell. Mol. Life Sci.* **67**, 179–199, https://doi.org/10.1007/s00018-009-0160-x
66 Yan, Y.-W., Mao, D.-D., Yang, L., Qi, J.-L., Zhang, X.-X., Tang, Q.-L. et al. (2018) Magnesium transporter MGT6 plays an essential role in maintaining magnesium homeostasis and regulating high magnesium tolerance in Arabidopsis. *Front. Plant Sci.* **9**, 274, https://doi.org/10.3389/fpls.2018.00274
67 Bugde, P., Biswas, R., Merien, F., Lu, J., Liu, D.-X., Chen, M. et al. (2017) The therapeutic potential of targeting ABC transporters to combat multi-drug resistance. *Expert Opin. Ther. Targets* **21**, 511–530, https://doi.org/10.1080/14728222.2017.1310841
68 Bull, P.C. and Cox, D.W. (1994) Wilson disease and Menkes disease: new handles on heavy-metal transport. *Trends Genet.* **10**, 246–252, https://doi.org/10.1016/0168-9525(94)90172-4
69 Bailey, S., Ward, D., Middleton, R., Grossmann, J.G. and Zambryski, P.C. (2006) Agrobacterium tumefaciens VirB8 structure reveals potential protein–protein interaction sites. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2582–2587, https://doi.org/10.1073/pnas.0511216103
70 Pierson, III, L.S., Gaffney, T., Lam, S. and Gong, F. (1995) Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium Pseudomonas aureofaciens 30-84. *FEMS Microbiol. Lett.* **134**, 299–307
71 Hunter, R.C., Klepac-Ceraj, V., Lorenzi, M.M., Grotzinger, H., Martin, T.R. and Newman, D.K. (2012) Phenazine content in the cystic fibrosis respiratory tract negatively correlates with lung function and microbial complexity. *Am. J. Respir. Cell Mol. Biol.* **47**, 738–745, https://doi.org/10.1165/rcmb.2012-0088OC
72 Upadhyay, A. and Srivastava, S. (2011) Phenazine-1-carboxylic acid is a more important contributor to biocontrol Fusarium oxysporum than pyrrolnitrin in Pseudomonas fluorescens strain Psd. *Microbiol. Res.* **166**, 323–335, https://doi.org/10.1016/j.micres.2010.06.001
73 Banco, M.T., Mishra, V., Greeley, S.C. and Ronning, D.R. (2018) Direct detection of products from S-adenosylmethionine-dependent enzymes using a competitive fluorescence polarization assay. *Anal. Chem.* **90**, 1740–1747, https://doi.org/10.1021/acs.analchem.7b03556
74 Martin, J.L. and McMillan, F.M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783–793, https://doi.org/10.1016/S0959-440X(02)00391-3
75 Ruan, J., Wu, X., Ye, Y. and Härdter, R. (1998) Effect of potassium, magnesium and sulphur applied in different forms of fertilisers on free amino acid content in leaves of tea (Camellia sinensis L). *J. Sci. Food Agric.* **76**, 389–396, https://doi.org/10.1002/(SICI)1097-0010(199803)76:3%3c389::AID-JSFA963%3e3.0.CO;2-X
76 Frias, J., Doblado, R., Antezana, J.R. and Vidal-Valverde, C. (2003) Inositol phosphate degradation by the action of phytase enzyme in legume seeds. *Food Chem.* **81**, 233–239, https://doi.org/10.1016/S0308-8146(02)00417-X
77 Dersjant-Li, Y., Awati, A., Schulze, H. and Partridge, G. (2015) Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. *J. Sci. Food Agric.* **95**, 878–896, https://doi.org/10.1002/jsfa.6998
78 Harada, J., Takaku, S. and Watanabe, K. (2012) An on-demand metalloprotease from psychro-tolerant Exiguobacterium undae Su-1, the activity and stability of which are controlled by the Ca2+ concentration. *Biosci. Biotechnol. Biochem.* **76**, 986–992, https://doi.org/10.1271/bbb.110997
79 Akoh, C.C., Lee, G.-C., Liaw, Y.-C., Huang, T.-H. and Shaw, J.-F. (2004) GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43**, 534–552, https://doi.org/10.1016/j.plipres.2004.09.002
80 KONG, K.F., Schneper, L. and Mathee, K. (2010) Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS* **118**, 1–36, https://doi.org/10.1111/j.1600-0463.2009.02563.x
81 Lam, W.W.L., Woo, E.J., Kotaka, M., Tam, W.K., Leung, Y.C., Ling, T.K.W. et al. (2010) Molecular interaction of flagellar export chaperone FliS and cochaperone HP1076 in Helicobacter pylori. *FASEB J.* **24**, 4020–4032, https://doi.org/10.1096/fj.10-155242
82 Meganathan, R. (2001) Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms. *Vitamins Hormones* **61**, 173–218
83 Debnath, J., Siricilla, S., Wan, B., Crick, D.C., Lenaerts, A.J., Franzblau, S.G. et al. (2012) Discovery of selective menaquinone biosynthesis inhibitors against Mycobacterium tuberculosis. *J. Med. Chem.* **55**, 3739–3755, https://doi.org/10.1021/jm201608g
84 Idicula-Thomas, S. and Balaji, P.V. (2005) Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. *Protein Sci.* **14**, 582–592, https://doi.org/10.1110/ps.041009005
85 Jaspard, E., Macherel, D. and Hunault, G. (2012) Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. *PLoS ONE* **7**, e36968, https://doi.org/10.1371/journal.pone.0036968
86 Meier, M., Sit, R.V. and Quake, S.R. (2013) Proteome-wide protein interaction measurements of bacterial proteins of unknown function. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 477–482, https://doi.org/10.1073/pnas.1210634110
87 Chance, M.R., Bresnick, A.R., Burley, S.K., Jiang, J.S., Lima, C.D., Sali, A. et al. (2002) Structural genomics: a pipeline for providing structures for the biologist. *Protein Sci.* **11**, 723–738, https://doi.org/10.1110/ps.4570102

88 Cardew, E.M., Verlinde, C.L. and Pohl, E. (2018) The calcium-dependent protein kinase 1 from Toxoplasma gondii as target for structure-based drug design. *Parasitology* **145**, 210–218, https://doi.org/10.1017/S0031182017001901

89 Kumar, A., Sharma, A., Kaur, G., Makkar, P. and Kaur, J. (2017) Functional characterization of hypothetical proteins of Mycobacterium tuberculosis with possible esterase/lipase signature: a cumulative in silico and in vitro approach. *J. Biomol. Struct. Dyn.* **35**, 1226–1243, https://doi.org/10.1080/07391102.2016.1174738

90 Choi, H.-P., Juarez, S., Ciordia, S., Fernandez, M., Bargiela, R., Albar, J.P. et al. (2013) Biochemical characterization of hypothetical proteins from Helicobacter pylori. *PLoS ONE.* **8**, https://doi.org/10.1371/journal.pone.0066605

91 Cort, J.R., Yee, A., Edwards, A.M., Arrowsmith, C.H. and Kennedy, M.A. (2000) NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from Methanobacterium thermoautotrophicum. *J. Struct. Funct. Genomics* **1**, 15–25, https://doi.org/10.1023/A:1011348803324

92 Barta, M.L., Thomas, K., Yuan, H., Lovell, S., Battaile, K.P., Schramm, V.L. et al. (2014) Structural and biochemical characterization of Chlamydia trachomatis hypothetical protein CT263 supports that menaquinone synthesis occurs through the futalosine pathway. *J. Biol. Chem.* **289**, 32214–32229, https://doi.org/10.1074/jbc.M114.594325

93 Zhang, W., Culley, D.E., Gritsenko, M.A., Moore, R.J., Nie, L., Scholten, J.C. et al. (2006) LC–MS/MS based proteomic analysis and functional inference of hypothetical proteins in Desulfovibrio vulgaris. *Biochem. Biophys. Res. Commun.* **349**, 1412–1419, https://doi.org/10.1016/j.bbrc.2006.09.019

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

37	WP_011812706.1	0	0	0	0	1	20
38	WP_009882097.1	0	0	0	0	0	0
39	WP_002859472.1	0	0	0	0	0	0
40	WP_011812708.1	0	0	0	0	0	0
41	WP_002857599.1	0	0	0	0	1	20
42	WP_009882129.1	1	0	0	0	0	20
43	WP_009882150.1	0	0	0	0	1	20
44	WP_002869238.1	0	0	0	0	0	0
45	WP_009882162.1	1	1	1	1	0	80
46	WP_002859498.1	0	0	0	0	0	0
47	WP_002857569.1	0	0	0	0	0	0
48	WP_002869241.1	0	0	0	0	0	0
49	WP_002869242.1	0	0	0	0	0	0
50	WP_002869268.1	0	0	0	0	0	0
51	WP_010790856.1	1	1	1	1	0	80
52	WP_002859017.1	0	0	0	0	0	0
53	WP_002868796.1	0	0	0	0	0	0
54	WP_002868795.1	0	0	0	0	0	0
55	WP_002859022.1	0	0	0	0	0	0
56	WP_002868785.1	0	0	0	0	0	0
57	WP_002868784.1	1	0	0	0	1	40
58	WP_009882239.1	1	1	1	1	0	80
59	WP_002869315.1	1	0	0	0	0	20
60	WP_002856797.1	0	0	0	0	0	0
61	WP_002855217.1	0	0	0	0	0	0
62	WP_002854991.1	0	1	1	1	0	60
63	WP_002776186.1	0	0	0	0	0	0
64	WP_010790923.1	0	0	0	0	0	0
65	WP_002869301.1	0	0	0	0	0	0
66	WP_011812719.1	1	0	0	0	0	20
67	WP_011812720.1	0	0	0	0	0	0
68	WP_002854816.1	0	0	0	0	0	0
69	WP_002869110.1	0	0	0	0	0	0
70	WP_002855063.1	0	0	0	0	0	0
71	WP_002855029.1	1	1	1	1	0	80
72	WP_024088100.1	1	0	0	0	0	20
73	WP_079005306.1	0	0	0	0	1	20
74	WP_002868898.1	0	0	0	0	0	0
75	WP_002868905.1	1	1	1	1	0	80
76	WP_002869362.1	0	0	0	0	0	0

77	WP_002869356.1	1	1	1	1	0	80
78	WP_002855032.1	0	0	0	0	0	0
79	WP_074468912.1	0	0	0	0	0	0
80	WP_002852327.1	0	0	0	0	0	0
81	WP_002854947.1	0	0	0	0	0	0
82	WP_002852223.1	0	0	0	0	0	0
83	WP_002852050.1	0	0	0	0	0	0
84	WP_002858514.1	1	0	0	0	0	20
85	WP_002856929.1	1	1	1	1	0	80
86	WP_002825403.1	0	0	0	0	0	0
87	WP_002869028.1	1	1	1	1	0	80
88	WP_002869375.1	0	0	0	0	0	0
89	WP_009882410.1	1	0	0	0	0	20
90	WP_009882413.1	1	0	0	0	0	20
91	WP_002869370.1	0	0	0	0	1	20
92	WP_009882434.1	1	0	0	0	0	20
93	WP_002852615.1	0	0	0	0	0	0
94	WP_002856776.1	0	0	0	0	0	0
95	WP_011812736.1	1	1	1	1	0	80
96	WP_009882443.1	0	0	0	0	0	0
97	WP_002867958.1	0	0	0	0	0	0
98	WP_009882473.1	0	0	0	0	0	0
99	WP_002856908.1	0	0	0	0	0	0
100	WP_002868809.1	1	1	1	1	1	100
101	WP_002868812.1	0	0	0	0	0	0
102	WP_002856855.1	0	0	0	0	0	0
103	WP_002868815.1	1	0	0	0	0	20
104	WP_011812741.1	0	0	0	0	0	0
105	WP_002852485.1	1	0	0	0	0	20
106	WP_002860516.1	1	0	0	0	0	20
107	WP_002858755.1	0	0	0	0	0	0
108	WP_002857162.1	0	0	0	0	0	0
109	WP_002856886.1	0	0	0	0	0	0
110	WP_011812746.1	0	0	0	0	0	0
111	WP_002869164.1	0	0	0	0	0	0
112	WP_002857165.1	0	0	0	0	0	0
113	WP_002856266.1	0	0	0	0	1	20
114	WP_002911887.1	0	0	0	0	0	0
115	WP_002856425.1	0	0	0	0	0	0
116	WP_011812749.1	0	0	0	0	0	0

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_002856369.1	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_002822505.1	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

157	WP_002854125.1	0	0	0	0	0	0
158	WP_002869326.1	1	1	1	1	0	80
159	WP_002855905.1	0	0	0	0	0	0
160	WP_002856036.1	0	0	0	0	0	0
161	WP_002859649.1	0	0	0	0	0	0
162	WP_002869147.1	1	0	0	0	0	20
163	WP_002869378.1	0	0	0	0	1	20
164	WP_002855633.1	0	0	0	0	0	0
165	WP_002862076.1	0	0	0	0	1	20
166	WP_002869139.1	1	1	1	1	0	80
167	WP_002860363.1	0	0	0	0	0	0
168	WP_002818956.1	1	0	0	0	0	20
169	WP_002867263.1	0	0	0	0	0	0
170	WP_004315270.1	0	0	0	0	1	20
171	WP_002855347.1	1	0	0	0	0	20
172	WP_002858858.1	0	0	0	0	0	0
173	WP_002869194.1	1	1	1	1	0	80
174	WP_002869195.1	0	1	1	1	0	60
175	WP_002855438.1	0	0	0	0	0	0
176	WP_002856630.1	1	1	1	1	0	80
177	WP_002869204.1	0	0	0	0	0	0
178	WP_002858872.1	0	0	0	0	0	0
179	WP_002855458.1	1	1	1	1	0	80
180	WP_002797496.1	1	1	1	1	0	80
181	WP_024088174.1	1	1	1	1	0	80
182	WP_009883030.1	1	1	1	1	0	80
183	WP_002922487.1	0	0	0	0	0	0
184	WP_002824979.1	1	1	1	1	0	80
185	WP_002869225.1	1	1	1	1	0	80
186	WP_002856158.1	0	0	0	0	0	0
187	WP_002934257.1	0	0	0	0	0	0
188	WP_002856602.1	1	1	1	1	0	80
189	WP_009883110.1	0	0	0	0	1	20
190	WP_002778908.1	0	0	0	0	0	0
191	WP_002868888.1	1	0	0	1	1	60
192	WP_002855727.1	0	0	0	0	0	0
193	WP_002868880.1	1	1	1	1	1	100
194	WP_009883121.1	1	1	1	1	0	80
195	WP_002868874.1	0	0	0	0	0	0
196	WP_002869278.1	0	0	0	0	0	0

197	WP_002860117.1	1	1	1	1	0	80
198	WP_079254198.1	0	0	0	0	0	0
199	WP_002882243.1	1	0	0	0	1	40
200	WP_002869298.1	0	0	0	0	0	0
201	WP_002853105.1	0	0	0	0	0	0
202	WP_002790440.1	1	0	0	0	0	20
203	WP_002790442.1	0	0	0	0	0	0
204	WP_002790713.1	0	0	0	0	0	0
205	WP_002779702.1	0	0	0	0	0	0
206	WP_002779703.1	0	0	0	0	0	0
207	WP_002779704.1	0	1	1	1	0	60
208	WP_002790730.1	1	0	0	0	0	20
209	WP_002804244.1	0	1	1	0	0	40
210	WP_011271766.1	0	0	0	0	0	0
211	WP_002779777.1	0	0	0	0	0	0
212	WP_002809140.1	0	0	0	0	0	0
213	WP_002826068.1	0	0	0	0	0	0
214	WP_002804272.1	0	0	0	0	0	0
215	WP_011187233.1	1	1	1	1	1	100
216	WP_011187234.1	1	0	0	0	1	40
217	WP_002844160.1	0	0	0	0	0	0
218	WP_011815226.1	0	0	0	0	0	0
219	WP_011187235.1	1	1	1	1	0	80
220	WP_002809051.1	1	0	0	0	0	20
221	WP_002809052.1	0	0	0	0	0	0
222	WP_002842869.1	0	0	0	0	0	0
223	WP_011187239.1	0	0	0	0	0	0
224	WP_032592775.1	0	0	0	0	1	20
225	WP_002809111.1	1	1	1	1	0	80
226	WP_002809110.1	0	0	0	0	0	0
227	WP_002809107.1	0	0	0	0	0	0
228	WP_002834241.1	0	0	0	0	0	0
229	WP_002779751.1	0	0	0	0	0	0
230	WP_002909884.1	1	0	0	0	0	20
231	WP_008976813.1	0	0	0	0	0	0
232	WP_002801797.1	0	0	0	0	0	0
233	WP_011117548.1	1	1	1	1	0	80
234	WP_011117549.1	1	1	1	1	0	80
235	WP_011117551.1	0	0	0	0	0	0
236	WP_002815556.1	0	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.

			PFAM		**SMART**		**MOTIF**		**INTERPROSCAN**		**CDART**		**SUPERFAMILY**		**SVMprot**	
No.	Protein ID	Protein Name_Known function	Prediction	Score	Prediction2	Score2	Prediction3	Score3	Prediction4	Score4	Prediction5	Score5	prediction6	Score6	prediction 7	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)	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)	glutamate synthase	1 (5)	Manganese-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)	Forming Carbon-Nitrogen Bonds, Transferases - Glycosyltransferases	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)	All lipid-binding proteins, Transferases - Glycosyltransferases	1 (2)
5	WP_009881534.1	transglycosylase	transglycosylase	1 (5)	transglycosylase	1 (5)	transglycosylase	1 (5)	transglycosylase	1 (5)	transglycosylase	1 (5)	transglycosylase	1 (5)	All DNA-binding, Magnesium-binding	1 (3)
6	WP_009881539.1	Bcr/CflA family efflux MFS transporter	Bcr/CflA family efflux MFS transporter	1 (5)	Bcr/CflA family efflux MFS transporter	1 (5)	Bcr/CflA family efflux MFS transporter	1 (5)	Bcr/CflA family efflux MFS transporter	1 (5)	Bcr/CflA family efflux MFS transporter	1 (5)	Bcr/CflA family efflux MFS transporter	1 (5)	Electrochemical Potential-driven transporters - Porters	1 (5)

7	WP_002854281.1	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_002859393.1	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)	acetylglutamate kinase	1 (5)
9	WP_002854336.1	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)	Electrochemical Potential-driven transporters	1 (4)
10	WP_002859852.1	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)	EamA-like transporter family	1 (4)	multidrug efflux SMR transporter	1 (5)	Metal-binding	1 (3)
11	WP_009882032.1	prephenate dehydratase	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_002858694.1	lysine--tRNA ligase	tRNA synthetases	1 (5)	tRNA synthetases	1 (5)	tRNA synthetases	1 (5)	tRNA synthetases	1 (5)	tRNA synthetases	1 (5)	tRNA synthetases	1 (5)	Forming Carbon-Oxygen Bonds	1 (3)
13	WP_009882169.1	YigZ family protein	UPF0029	1 (2)	UPF0029	1 (2)	UPF0029	1 (2)	Impact family	1 (2)	UPF0029	1 (2)	YigZ family protein	1 (5)	Transferring Phosphorus-Containing Groups	1 (2)
14	WP_002857383.1	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_002782934.1	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)	Nucleic acid-binding proteins	0 (2)	rRNA-binding proteins	1 (4)
16	WP_002857290.1	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)

															Phosphorus-Oxygen Lyases	
17	WP_002869243.1	ferrochelatase	ferrochelatase	1 (5)	ferrochelatase	1 (5)	ferrochelatase	1 (5)	ferrochelatase	1 (5)	ferrochelatase	1 (5)	ferrochelatase	1 (5)	Transferring Phosphorus-Containing Groups, All lipid-binding proteins	1 (3)
18	WP_002854879.1	flagellar basal body rod protein FlgB	No result	0 (2)	No result	0 (2)	No result	0 (2)	flagellar basal body rod protein FlgB	1 (5)	flagellar basal body rod protein FlgB	1 (5)	No result	0 (2)	Structural proteins	1 (4)
19	WP_002880964.1	endolytic transglycosylase MltG	endolytic transglycosylase MltG	1 (5)	endolytic transglycosylase MltG	1 (5)	endolytic transglycosylase MltG	1 (5)	endolytic transglycosylase MltG	1 (5)	endolytic transglycosylase MltG	1 (5)	No result	0 (2)	Transferases - Glycosyltransferases	1 (4)
20	WP_002856958.1	ATP-binding cassette domain-containing protein	ATP-binding cassette domain-containing protein	1 (5)	ATP-binding cassette domain-containing protein	1 (5)	ATP-binding cassette domain-containing protein	1 (5)	ATP-binding cassette domain-containing protein	1 (5)	ATP-binding cassette domain-containing protein	1 (5)	ATP-binding cassette domain-containing protein	1 (5)	Acting on Acid Anhydrides	1 (3)
21	WP_002868904.1	TolC family protein	TolC family protein	1 (5)	TolC family protein	1 (5)	TolC family protein	1 (5)	TolC family protein	1 (5)	TolC family protein	1 (5)	TolC family protein	1 (5)	Zinc-binding	0 (2)
22	WP_002869361.1	carbamoyltransferase HypF	carbamoyltransferase HypF	1 (5)	carbamoyltransferase HypF	1 (5)	carbamoyltransferase HypF	1 (5)	carbamoyltransferase HypF	1 (5)	carbamoyltransferase HypF	1 (5)	carbamoyltransferase HypF	1 (5)	Zinc-binding	0 (2)
23	WP_002869360.1	hydrogenase formation protein HypD	hydrogenase formation protein HypD	1 (5)	hydrogenase formation protein HypD	1 (5)	hydrogenase formation protein HypD	1 (5)	hydrogenase formation protein HypD	1 (5)	hydrogenase formation protein HypD	1 (5)	No result	0 (2)	Zinc-binding, Carbon-Carbon Lyases	1 (3)
24	WP_002869354.1	aspartate--tRNA ligase	aspartate--tRNA ligase	1 (5)	aspartate--tRNA ligase	1 (5)	aspartate--tRNA ligase	1 (5)	aspartate--tRNA ligase	1 (5)	aspartate--tRNA ligase	1 (5)	aspartate--tRNA ligase	1 (5)	Forming Carbon-Oxygen Bonds, Zinc-binding	1 (3)
25	WP_002869349.1	MFS transporter	Proton-dependent	1 (3)	MFS transporter	1 (5)	MFS transporter	1 (5)	MFS transporter	1 (5)	MFS transporter	1 (5)	MFS transporter	1 (5)	Iron-binding	1 (3)

			oligopeptide transporter												
26	WP_002869372.1	molecular chaperone DnaK	molecular chaperone DnaK	1 (5)	molecular chaperone DnaK	1 (5)	molecular chaperone DnaK	1 (5)	molecular chaperone DnaK	1 (5)	molecular chaperone DnaK	1 (5)	molecular chaperone DnaK	1 (5)	All lipid-binding proteins	1 (3)
27	WP_009882420.1	nucleotide exchange factor GrpE	nucleotide exchange factor GrpE	1 (5)	nucleotide exchange factor GrpE	1 (5)	nucleotide exchange factor GrpE	1 (5)	nucleotide exchange factor GrpE	1 (5)	nucleotide exchange factor GrpE	1 (5)	nucleotide exchange factor GrpE	1 (5)	Zinc-binding, Acting on Acid Anhydrides	1 (3)
28	WP_002857174.1	serine O-acetyltransferase	Hexapeptide repeat	1 (2)	Hexapeptide repeat	1 (2)	Hexapeptide repeat	1 (2)	serine O-acetyltransferase	1 (5)	Carbonic anhydrase or acetyltransferase, isoleucine patch superfamily	0 (2)	serine O-acetyltransferase	1 (5)	Phosphotransfer-driven group translocators , Manganese-binding	0 (2)
29	WP_011812734.1	ATP-dependent helicase	ATP-dependent helicase	1 (5)	ATP-dependent helicase	1 (5)	ATP-dependent helicase	1 (5)	ATP-dependent helicase	1 (5)	ATP-dependent helicase	1 (5)	ATP-dependent helicase	1 (5)	Zinc-binding, All DNA-binding	1 (3)
30	WP_002867950.1	alpha/beta hydrolase	alpha/beta hydrolase	1 (5)	alpha/beta hydrolase	1 (5)	alpha/beta hydrolase	1 (5)	alpha/beta hydrolase	1 (5)	alpha/beta hydrolase	1 (5)	alpha/beta hydrolase	1 (5)	All lipid-binding proteins	1 (3)
31	WP_002869103.1	c-type cytochrome	c-type cytochrome	1 (5)	c-type cytochrome	1 (5)	c-type cytochrome	1 (5)	c-type cytochrome	1 (5)	c-type cytochrome	1 (5)	c-type cytochrome	1 (5)	Zinc-binding	1 (3)
32	WP_011812744.1	DNA translocase FtsK	DNA translocase FtsK	1 (5)	DNA translocase FtsK	1 (5)	DNA translocase FtsK	1 (5)	DNA translocase FtsK	1 (5)	DNA translocase FtsK	1 (5)	DNA translocase FtsK	1 (5)	Zinc-binding, Acting on peptide bonds (Peptidases)	1 (3)
33	WP_002853404.1	GNAT family N-acetyltransferase	GNAT family N-acetyltransferase	1 (5)	GNAT family N-acetyltransferase	1 (5)	GNAT family N-acetyltransferase	1 (5)	GNAT family N-acetyltransferase	1 (5)	GNAT family N-acetyltransferase	1 (5)	GNAT family N-acetyltransferase	1 (5)	All lipid-binding proteins	1 (3)
34	WP_002853451.1	RNA polymerase sigma factor	RNA polymerase sigma factor	1 (5)	RNA polymerase sigma factor	1 (5)	RNA polymerase sigma factor	1 (5)	RNA polymerase sigma factor	1 (5)	RNA polymerase sigma factor RpoD	1 (5)	RNA polymerase sigma factor	1 (5)	DNA-directed RNA polymeras	1 (5)

		RpoD	RpoD		RpoD		RpoD		RpoD				RpoD		e	
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)	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)	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)	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)	riboflavin synthase	1 (5)	riboflavin synthase	1 (5)	Transferring Alkyl or Aryl Groups, Other than Methyl Groups	1 (3)
39	WP_002855731.1	bacteriohemerythrin	bacteriohemerythrin	1 (5)	bacteriohemerythrin	1 (5)	bacteriohemerythrin	1 (5)	bacteriohemerythrin	1 (5)	bacteriohemerythrin	1 (5)	bacteriohemerythrin	1 (5)	Zinc-binding, All DNA-binding	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)	GDP-L-fucose synthase	1 (5)	Short-chain dehydrogenases/reductases (SDR)	0 (2)	AD(P)-binding Rossmann-fold domains	0 (2)	server error	0 (2)

S4 Table. Annotation dataset results for the 50 HPs submitted to the workflow with Pfam, SMART, MOTIF, INTERPROSCAN, CDART, SUPERFAMILY and SVMprot

No.	Protein ID	PFAM	SMART	MOTIF	INTERPROSCAN	CDART	SUPERFAMILY	SVMProt
1	WP_002868767.1	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_002854524.1	Chemotaxis phosphatase CheX	CheX	Chemotaxis phosphatase CheX	chemotaxiz phosphatase cheX- like domain	CheC-like family	CheC-like	Transferring Phosphorus-Containing Groups
3	WP_009882162.1	SprA-related family	SprA-related	SprA-related family	SprA-related family	SprA-related family	No result	Zinc-binding
4	WP_010790856.1	Pyridoxamine 5'-phosphate oxidase	Pyridox_oxidase	Pyridoxamine 5'-phosphate oxidase	Pyridoxamine 5'-phosphate oxidase	Pyridoxine 5'-phosphate (PNP) oxidase-like	PNP-oxidase like	Zinc-binding
5	WP_009882239.1	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_002854991.1	FxsA cytoplasmic membrane protein, FxsA	FxsA	FxsA cytoplasmic membrane protein	FxsA cytoplasmic membrane protein	No result	No result	TC1.E Channels
7	WP_002855029.1	DNA replication regulator, HobA	HobA	DNA replication regulator	DNA replication regulator, HobA	DNA replication regulator, HobA Superfamily	SMI1	All lipid-binding proteins
8	WP_002868905.1	GDSL-like Lipase	Lipase_GDSL	GDSL-like Lipase	GDSL lipase	SGNH_hydrolase Superfamily/esterases and lipases	SGNH hydrolase/esterases and lipases	Zinc-binding
9	WP_002869356.1	Divergent polysaccharide deacetylase	Polysacc_deac_2	Divergent polysaccharide deacetylase	Divergent polysaccharide deacetylase	Divergent polysaccharide deacetylase	Divergent polysaccharide deacetylase	EC3.1 Hydrolases - Acting on Ester Bonds
10	WP_002856929.1	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_002869028.1	Esterase-like activity of phytase	Phytase-like	Esterase-like activity of phytase	Phytase-like domain	SdiA-regulated Superfamily	No result	Zinc-binding
12	WP_011812736.1	DUF234	DUF234	DUF234	Domain of unknown function DUF234	DUF4143 Superfamily	Restriction endonuclease-like	Zinc-binding

13	WP_002868809.1	Ankyrin repeats, Ank_2	Ankyrin repeats, Ank_2	Ankyrin repeats, Ank_2	Ankyrin repeat-containing domain	ANK Superfamily	Ankyrin repeat	EC3.2 Hydrolases - Glycosylases
14	WP_002869368.1	Type-1V conjugative transfer system mating pair stabilisation, TraN	TraN	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 Superfamily	TB module	All lipid-binding proteins
15	WP_009882583.1	NLPC_P60	NLPC_P60	NLPC_P60 stabilising domain, N term	NLPC/P60, N-terminal domain	NlpC/P60 family	NlpC/P60	Forming Carbon-Oxygen Bonds
16	WP_002853389.1	Jag N-terminus	Jag_N	Jag N-terminus	Jag, N-terminal domain superfamily	Jag N-terminus	No result	Transferring One-Carbon Groups
17	WP_009882608.1	Adhesin from Campylobacter	JLPA	Adhesin from Campylobacter	Adhesin JlpA, Campylobacter	JLPA Superfamily , Adhesin from Campylobacter	No result	Zinc-binding
18	WP_002856369.1	Putative beta-lactamase-inhibitor-like	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_079254190.1	Beta-1,4-N-acetylgalactosaminyltransferase (CgtA)	CgtA	Beta-1,4-N-acetylgalactosaminyltransferase (CgtA)	Beta-1,4-N-acetylgalactosaminyltransferase	Beta-1,4-N-acetylgalactosaminyltransferase (CgtA)	No result	Magnesium-binding
20	WP_002856180.1	No result	HMA	Heavy-metal-associated domain	Heavy metal-associated domain	Heavy-metal-associated domain (HMA)	HMA, heavy metal-associated domain	Copper-binding
21	WP_002831611.1	Transcription factor zinc-finger	zf-TFIIB	Transcription factor zinc-finger	Transcription factor zinc-finger	Transcription factor zinc-finger	No result	Magnesium-binding
22	WP_002790076.1	Methyl-accepting chemotaxis protein (MCP) signalling domain	Methyl-accepting chemotaxis-like domains	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) signaling domain	P-P-bond-hydrolysis-driven transporters
23	WP_002853792.1	Plasminogen-binding protein pgbA N-terminal	PGBA_N	Plasminogen-binding protein pgbA N-terminal	Plasminogen-binding protein PgbA, N-terminal	Plasminogen-binding protein pgbA N-terminal	No result	Manganese-binding
24	WP_002869072.1	Putative S-adenosyl-L-methionine-dependent methyltransferase	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_002869097.1	MaoC like domain	MaoC_dehydratas	MaoC like domain	MaoC-like dehydratase domain	Short-chain dehydrogenases	MaoC-like	Transferring Phosphorus-

								Containing Groups
26	WP_002869326.1	Carboxypeptidase controlling helical cell shape catalytic	Peptidase_M99	Carboxypeptidase controlling helical cell shape catalytic	Metallo-carboxypeptidase, C-terminal domain	C-terminal domain of metallo-carboxypeptidase	Zn-dependent exopeptidases	Zinc-binding
27	WP_002869139.1	Pyruvate phosphate dikinase, PEP	PPDK_N	Pyruvate phosphate dikinase, PEP	Pyruvate phosphate dikinase, PEP	Pyruvate phosphate dikinase, PEP	Pyruvate phosphate dikinase, N-terminal domain	Forming Carbon-Oxygen Bonds
28	WP_002869194.1	No result	Rod-binding	Rod binding protein	Uncharacterised conserved protein UCP007248	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase	No result	Acting on peptide bonds (Peptidases)
29	WP_002869195.1	Anti-sigma-28 factor	FlgM	Anti-sigma-28 factor	Anti-sigma-28 factor FlgM superfamily	No result	Anti-sigma factor FlgM	Sodium-binding
30	WP_002856630.1	PD-(D/E)XK nuclease superfamily	PDDEXK_1	PD-(D/E)XK nuclease superfamily, PDDEXK_1	PD-(D/E)XK endonuclease-like domain, AddAB-type	Inactivated superfamily I helicase	P-loop containing nucleoside triphosphate hydrolases	Forming Carbon-Oxygen Bonds
31	WP_002855458.1	MgtE intracellular N domain	MgtE_N	MgtE intracellular N domain	Magnesium transporter, MgtE intracellular domain	Flagellar motility protein MotE, a chaperone for MotC folding	MgtE N-terminal domain-like	ATP-binding cassette (ABC) family
32	WP_002797496.1	Flagellar FliJ protein	FliJ	Flagellar FliJ protein	Flagellar export FliJ	Flagellar FliJ protein	No result	Magnesium-binding
33	WP_024088174.1	Nitrate reductase delta subunit	Nitrate_red_del	Nitrate reductase delta subunit	Nitrate reductase chaperone	Nitrate reductase delta subunit	TorD-like	Type II (general) secretory pathway (IISP) family
34	WP_009883030.1	AAA domain, putative AbiEii toxin, Type IV TA system, AAA_21	AAA_21	AAA domain, putative AbiEii toxin, Type IV TA system, AAA_21	ATPase, AAA-type, core	AAA domain, putative AbiEii toxin, Type IV TA system, AAA_21 Superfamily	ABC transporter ATPase domain-like	Glycosyltransferases
35	WP_002824979.1	putative NADH-ubiquinone oxidoreductase chain E	NADH-UOR_E	putative NADH-ubiquinone oxidoreductase chain E	Putative NADH-ubiquinone oxidoreductase chain E	putative NADH-ubiquinone oxidoreductase chain E	SirA-like	All lipid-binding proteins
36	WP_002869225.1	DMSO reductase anchor subunit (DmsC)	DmsC	DMSO reductase anchor subunit (DmsC)	DMSO reductase anchor subunit (DmsC)	DMSOR_beta-like Superfamily	No result	Acting on a heme group of donors
37	WP_002856602.1	Putative beta-lactamase-inhibitor-like	PepSY_like	Putative beta-lactamase-inhibitor-like	Putative beta-lactamase-inhibitor-like, PepSY-like	Putative beta-lactamase-inhibitor-like, PepSY-like	BT0923-like	All lipid-binding proteins

38	WP_002868888.1	No result	TPR_2/ TPR_8	Tetratricopeptide repeat, TPR_2	Tetratricopeptide repeat	Lipopolysaccharide biosynthesis regulator YciM, contains six TPR domains and a predicted metal-binding C-terminal domain	Tetratricopeptide repeat (TPR)	Zinc-binding
39	WP_002868880.1	ABC-type transport auxiliary lipoprotein component	ABC_trans_aux	ABC-type transport auxiliary lipoprotein component	ABC-type transport auxiliary lipoprotein component	ABC-type transport auxiliary lipoprotein component	XCC0632-like	Carbon-Oxygen Lyases
40	WP_009883121.1	Flagellar FLiS export co-chaperone	FliS_cochap	Flagellar FLiS export co-chaperone, HP1076	Flagellar FLiS export co-chaperone, HP1076	Flagellar FLiS export co-chaperone, HP1076	No result	All DNA-binding
41	WP_002860117.1	Menaquinone biosynthesis	VitK2_biosynth	Menaquinone biosynthesis	Menaquinone biosynthesis enzyme	member of the type 2 periplasmic binding fold protein superfamily	Periplasmic binding protein-like II	Glycosyltransferases
42	WP_002779704.1	T-antigen specific domain	Papo_T_antigen	T-antigen specific domain	Small/middle T-antigen superfamily	No result	T-antigen specific domain-like	Acting on iron-sulfur proteins as donors
43	WP_011187233.1	Toprim domain	TOPRIM	Toprim domain	twinkle, TOPRIM domain	Uncharacterized domain associated with phage	No result	Zinc-binding
44	WP_011187235.1	AAA domain, AAA_25	AAA_25	AAA domain, AAA_25	P-loop containing nucleoside triphosphate hydrolase	TOPRIM Superfamily	P-loop containing nucleoside triphosphate hydrolases	Acting on Ester Bonds
45	WP_002809111.1	TrbM	TrbM	TrbM	TrbM	TrbM Superfamily	No result	Acting on a sulfur group of donors
46	WP_011117548.1	VirB8 protein	VirB8	VirB8 protein	Bacterial virulence protein VirB8	Virulence protein VirB8	VirB8-like	Metal-binding
47	WP_011117549.1	Conjugal transfer protein	CagX	Conjugal transfer protein	CagX	CagX	No result	Glycosyltransferases
48	WP_011117575.1	Type IV secretion system proteins,T4SS	T4SS	Type IV secretion system proteins, T4SS	Type IV secretion system, VirB5	VirB5 protein family	Type IV secretion system protein TraC	Type IV secretion system protein TraC
49	WP_011799393.1	TrbM	TrbM	TrbM	TrbM	TrbM	No result	Acting on Ester Bonds
50	WP_011117588.1	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 PemK-like	PemK-like, MazF-like toxin of type II toxin-antitoxin system	PemK	All DNA-binding

S5 Table. Results of the blastp search for similar sequences against non-reduntant (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-CJGB95377	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
2	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*	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-CJGB96114	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

6	WP_002854991.1	*Campylobacter jejuni RM1221*	100%	8.00E-83	247	99%	integral membrane protein
		Campylobacter jejuni subsp. jejuni 260.94	100%	1.00E-82	247	99%	integral membrane protein
		Campylobacter jejuni subsp. jejuni HB93-13	100%	2.00E-82	246	99%	integral membrane protein
		Campylobacter jejuni subsp. jejuni D2600	100%	2.00E-82	246	99%	integral membrane protein
		Campylobacter jejuni subsp. jejuni LMG 23264	100%	2.00E-82	246	98%	integral membrane protein
7	WP_002855029.1	*Campylobacter jejuni subsp. jejuni 260.94*	100%	2.00E-124	357	99%	hypothetical protein
		Campylobacter jejuni subsp. jejuni ICDCCJ07001	100%	6.00E-124	355	99%	hypothetical protein
		Campylobacter jejuni subsp. jejuni 129-258	100%	6.00E-124	355	99%	hypothetical protein
		Campylobacter jejuni K1	100%	6.00E-124	355	98%	hypothetical protein
8	WP_002868905.1	*Campylobacter jejuni subsp. jejuni 129-258*	100%	0.0	787	100%	hypothetical protein
		Campylobacter jejuni subsp. jejuni 81-176-DRH212	100%	0.0	786	99%	hypothetical protein
		Campylobacter jejuni subsp. jejuni CG8486	100%	0.0	784	99%	hypothetical protein
9	WP_002869356.1	*Campylobacter jejuni subsp. jejuni 129-258*	100%	0.0	718	99%	polysaccharide deacetylase
		Campylobacter jejuni subsp. jejuni 2008-872	100%	0.0	716	99%	polysaccharide deacetylase
		Campylobacter jejuni subsp. jejuni 1997-11	100%	0.0	716	99%	polysaccharide deacetylase
		Campylobacter jejuni subsp. jejuni 1798	100%	0.0	716	99%	polysaccharide deacetylase
		Campylobacter jejuni K1	100%	0.0	716	99%	polysaccharide deacetylase
10	WP_002856929.1	*Campylobacter jejuni RM1221*	100%	3.00E-167	470	100%	zinc ribbon domain protein
		Campylobacter jejuni subsp. jejuni S3	100%	1.00E-166	468	99%	zinc ribbon domain protein
		Campylobacter jejuni subsp. jejuni 327	100%	1.00E-166	468	99%	zinc ribbon domain protein
		Campylobacter jejuni CVM 41910	100%	1.00E-166	468	99%	zinc ribbon domain protein
		Campylobacter jejuni CVM 41927	100%	2.00E-166	468	99%	zinc ribbon domain protein
11	WP_002869028.1	*Campylobacter jejuni subsp. jejuni 81-*	100%	0.0	884	100%	glycerophosphodiester

		176-DRH212					phosphodiesterase
		Campylobacter jejuni subsp. jejuni 81-176-UMCW7	100%	0.0	883	99%	glycerophosphodiester phosphodiesterase
		Campylobacter sp. BCW_4319	100%	0.0	882	99%	glycerophosphodiester phosphodiesterase
		Campylobacter sp. 1	100%	0.0	882	99%	glycerophosphodiester phosphodiesterase
		Campylobacter jejuni 30318	100%	0.0	881	99%	glycerophosphodiester 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
		Campylobacter jejuni subsp. jejuni ICDCCJ07001	100%	0.0	563	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

		Campylobacter sp. 1	100%	0.0	902	99%	SH3_6 and SH3_7 domain-containing protein
		Campylobacter jejuni subsp. jejuni 51494	100%	0.0	902	99%	SH3_6 and SH3_7 domain-containing protein
16	WP_002853389.1	*Campylobacter jejuni subsp. jejuni CF93-6*	100%	0.0	551	99%	RNA-binding protein
		Campylobacter jejuni subsp. jejuni 84-25	100%	0.0	550	99%	RNA-binding protein
		Campylobacter jejuni subsp. jejuni 1336	100%	0.0	550	99%	RNA-binding protein
		Campylobacter jejuni subsp. jejuni IA3902	100%	0.0	550	99%	RNA-binding protein
		Campylobacter jejuni subsp. jejuni 305	100%	0.0	550	99%	RNA-binding protein
17	WP_009882608.1	*Campylobacter jejuni subsp. jejuni 81-176-UMCW7*	100%	0.0	727	99%	lipoprotein
		Campylobacter jejuni subsp. jejuni 81-176-UMCW9	100%	0.0	727	99%	lipoprotein
		Campylobacter sp. BCW_4319	100%	0.0	725	99%	lipoprotein
		Campylobacter jejuni subsp. jejuni HB93-13	100%	0.0	724	99%	lipoprotein
		Campylobacter jejuni subsp. jejuni 2008-894	100%	0.0	724	99%	lipoprotein
18	WP_002856369.1	*Campylobacter jejuni subsp. jejuni 260.94*	100%	3.00E-92	272	99%	Putative beta-lactamase-inhibitor-like, PepSY-like
		Campylobacter jejuni subsp. jejuni 81116	100%	1.00E-91	271	99%	Putative beta-lactamase-inhibitor-like, PepSY-like
		Campylobacter jejuni subsp. jejuni M1	100%	2.00E-91	270	99%	Putative beta-lactamase-inhibitor-like, PepSY-like
		Campylobacter jejuni subsp. jejuni ICDCCJ07001	100%	3.00E-91	270	99%	Putative beta-lactamase-inhibitor-like, PepSY-like
		Campylobacter jejuni subsp. jejuni 327	100%	6.00E-91	269	99%	Putative beta-lactamase-inhibitor-like, PepSY-like
19	WP_079254190.1	*Campylobacter jejuni subsp. jejuni 81-176*	100%	8.00E-31	110	100%	hypothetical protein
		Campylobacter sp. US54	100.00%	8.00E-31	110	100%	hypothetical protein
20	WP_002856180.1	*Campylobacter jejuni subsp. jejuni 81-176-DRH212*	100%	6.00E-36	124	100%	heavy-metal-associated domain
		Campylobacter jejuni subsp. jejuni 81-176-UMCW7	100%	2.00E-35	123	98.44%	heavy-metal-associated domain

		Campylobacter jejuni CVM 41974	100%	3.00E-35	122	96.88%	heavy-metal-associated domain
		Campylobacter sp. BCW_4319	100%	8.00E-35	121	96.88%	heavy-metal-associated domain
		Campylobacter sp. CH278	100%	9.00E-35	121	96.88%	heavy-metal-associated domain
21	WP_002831611.1	*Campylobacter jejuni subsp. jejuni LMG 23211*	100%	2.00E-59	185	100%	hypothetical protein
		Campylobacter jejuni subsp. jejuni LMG 9872	100%	3.00E-59	185	98.85%	hypothetical protein
		Campylobacter sp. CH186	100%	3.00E-59	185	98.85%	hypothetical protein
		Campylobacter sp. US55	100%	6.00E-59	184	98.85%	hypothetical protein
		Campylobacter sp. US53	100%	7.00E-59	184	98.85%	hypothetical protein
22	WP_002790076.1	*Campylobacter jejuni subsp. jejuni 81-176*	100%	0.0	919	100.00%	methyl-accepting chemotaxis protein
		Campylobacter jejuni subsp. jejuni LMG 9879	100%	0.0	918	99.56%	methyl-accepting chemotaxis protein
		Campylobacter jejuni subsp. jejuni 2008-1025	100%	0.0	918	99.78%	methyl-accepting chemotaxis protein
		Campylobacter jejuni subsp. jejuni 260.94	100%	0.0	918	99.78%	methyl-accepting chemotaxis protein
23	WP_002853792.1	*Campylobacter jejuni CVM 41973*	100%	1.00E-173	486	100.00%	exporting protein
		Campylobacter jejuni CVM 41910	100%	2.00E-173	486	99.59%	exporting protein
		Campylobacter jejuni CVM 41922	100%	2.00E-173	486	99.59%	exporting protein
		Campylobacter jejuni CVM 41914	100%	3.00E-173	485	99.59%	exporting protein
		Campylobacter jejuni CVM 41936	100%	6.00E-173	484	99.59%	exporting protein
24	WP_002869072.1	*Campylobacter sp. BCW_4319*	100%	0.0	629	100.00%	hypothetical protein
		Campylobacter jejuni subsp. jejuni str. RM3420	100%	0.0	627	99.68%	hypothetical protein
		Campylobacter jejuni BJ-CJD101	100%	0.0	627	99.68%	hypothetical protein
		Campylobacter jejuni CVM 41974	100%	0.0	623	99.37%	hypothetical protein
		Campylobacter jejuni subsp. jejuni 140-16	100%	0.0	621	98.10%	hypothetical protein
25	WP_002869097.1	*Campylobacter jejuni subsp. jejuni 81-176*	100%	0.0	895	100.00%	MaoC like domain
		Campylobacter sp. BCW_4319	100%	0.0	892	99.78%	MaoC like domain
		Campylobacter sp. US54	100%	0.0	892	99.78%	MaoC like domain

		Campylobacter coli LMG 23342	100%	0.0	891	99.34%	MaoC like domain
		Campylobacter coli 15-537360	100%	0.0	890	99.56%	MaoC like domain
26	WP_002869326.1	*Campylobacter jejuni CVM 41974*	100%	0.0	945	100.00%	deacylase
		Campylobacter jejuni subsp. jejuni	100%	0.0	940	99.35%	deacylase
		Campylobacter sp. US12a	100%	0.0	940	99.35%	deacylase
		Campylobacter sp. CH165	100%	0.0	939	99.35%	deacylase
		Campylobacter sp. CH278	100%	0.0	939	99.35%	deacylase
27	WP_002869139.1	*Campylobacter jejuni subsp. jejuni 129-258*	100%	0.0	1579	100.00%	hypothetical protein
		Campylobacter sp. US54	100%	0.0	1576	99.74%	hypothetical protein
		Campylobacter jejuni 30318	100%	0.0	1573	99.61%	hypothetical protein
		Campylobacter jejuni subsp. jejuni 53161	100%	0.0	1572	99.49%	hypothetical protein
		Campylobacter jejuni subsp. jejuni 87459	100%	0.0	1571	99.49%	hypothetical protein
28	WP_002869195.1	*Campylobacter jejuni subsp. jejuni 129-258*	100%	5.00E-37	127	100.00%	flagellar biosynthesis anti-sigma factor FlgM
		Campylobacter sp. US54	100%	1.00E-36	126	98.46%	flagellar biosynthesis anti-sigma factor FlgM
		Campylobacter jejuni subsp. jejuni M129	100%	2.00E-36	125	98.46%	flagellar biosynthesis anti-sigma factor FlgM
		Campylobacter jejuni subsp. jejuni 1336	100%	3.00E-36	125	96.92%	flagellar biosynthesis anti-sigma factor FlgM
		Campylobacter coli 2680	100%	5.00E-36	124	96.92%	flagellar biosynthesis anti-sigma factor FlgM
29	WP_002856630.1	*Campylobacter jejuni subsp. jejuni str. RM3420*	100%	0.0	1560	100.00%	helicase AddB
		Campylobacter jejuni K5	100%	0.0	1559	99.87%	helicase AddB
		Campylobacter jejuni subsp. jejuni 1577	100%	0.0	1557	99.87%	helicase AddB
30	WP_002855458.1	*Campylobacter jejuni RM1221*	100%	6.00E-117	337	100.00%	nucleosidase
		Campylobacter jejuni subsp. jejuni S3	100%	9.00E-117	337	99.42%	nucleosidase
		Campylobacter jejuni Cj2	100%	2.00E-116	336	99.42%	nucleosidase
		Campylobacter jejuni 20176	100%	2.00E-116	336	99.42%	nucleosidase

		Campylobacter jejuni CVM 41922	100%	2.00E-116	336	99.42%	nucleosidase
31	WP_002797496.1	*Campylobacter jejuni RM1221*	100%	2.00E-95	280	100.00%	hypothetical protein
		Campylobacter coli 15-537360	100%	6.00E-95	279	99.30%	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
32	WP_024088174.1	*Campylobacter jejuni subsp. jejuni M129*	100%	6.00E-167	469	100.00%	formate dehydrogenase-specific chaperone
		Campylobacter jejuni subsp. jejuni IA3902	100%	2.00E-166	468	99.58%	formate dehydrogenase-specific chaperone
		Campylobacter jejuni subsp. jejuni S3	100%	3.00E-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 sp. 112	100%	5.00E-166	467	99.58%	formate dehydrogenase-specific chaperone
33	WP_009883030.1	*Campylobacter jejuni subsp. jejuni 81-176-UMCW7*	100%	0.0	865	100.00%	ATP/GTP-binding protein
		Campylobacter jejuni subsp. jejuni 81-176	100%	0.0	865	100.00%	ATP/GTP-binding protein
		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
		Streptococcus phocae	100%	0.0	862	99.77%	ATP/GTP-binding protein
34	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 RM4661	100%	9.00E-45	147	98.67%	NADH-ubiquinone oxidoreductase
35	WP_002869225.1	*Campylobacter sp. BCW_4319*	100%	0.0	577	100.00%	dimethylsulfoxide reductase
		Campylobacter jejuni subsp. jejuni 81-176	100%	0.0	575	99.65%	dimethylsulfoxide reductase
		Campylobacter jejuni subsp. jejuni str. RM3420	100%	0.0	574	99.65%	dimethylsulfoxide reductase

		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 ICDCCJ07001	100%	3.00E-94	277	99.28%	periplasmic protein
		Campylobacter jejuni subsp. jejuni str. RM3420	100%	5.00E-94	276	99.28%	periplasmic protein
		Campylobacter jejuni subsp. jejuni 1336	100%	5.00E-94	276	98.55%	periplasmic protein
37	WP_002868888.1	*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

		Campylobacter jejuni CVM 41900	100%	0.0	582	99.65%	S-ribosylhomocysteine lyase
		Campylobacter jejuni subsp. jejuni 1997-1	100%	0.0	582	99.65%	S-ribosylhomocysteine lyase
		Campylobacter sp. CH246	100%	0.0	582	99.65%	S-ribosylhomocysteine lyase
		Campylobacter sp. BCW_6461	100%	0.0	582	99.65%	S-ribosylhomocysteine lyase
41	WP_002779704.1	*Campylobacter jejuni subsp. jejuni 81-176*	100%	5.00E-56	176	100.00%	cpp11 like protein
		Campylobacter jejuni subsp. jejuni S3	100%	8.00E-56	176	100.00%	cpp11 like protein
		Campylobacter coli CVM N29710	100%	3.00E-55	175	98.86%	cpp11 like protein
		Campylobacter coli 2553	98%	6.00E-55	174	100.00%	cpp11 like protein
		Campylobacter coli 86119	100%	1.00E-54	173	97.73%	cpp11 like protein
42	WP_011187233.1	*Campylobacter jejuni subsp. jejuni 81-176*	100%	0.0	828	100%	cpp22 like protein
		Campylobacter jejuni subsp. jejuni 81-176	100%	0.0	827	99.75%	cpp22 like protein
		Campylobacter coli 1148	100%	0.0	827	99.75%	cpp22 like protein
		Campylobacter jejuni CVM 41946	100%	0.0	826	99.75%	cpp22 like protein
		Campylobacter sp. BCW_6462	100%	0.0	826	99.51%	cpp22 like protein
43	WP_011187235.1	*Campylobacter jejuni subsp. jejuni 81-176*	100%	0.0	1196	100.00%	cpp26 like protein
		Campylobacter jejuni subsp. jejuni M129	100%	0.0	1195	99.83%	cpp26 like protein
		Campylobacter jejuni subsp. jejuni D42a	100%	0.0	1193	99.83%	cpp26 like protein
44	WP_002809111.1	Campylobacter jejuni subsp. jejuni 81-176	100%	0.0	524	100.00%	conjugal transfer protein TrbM
		Campylobacter coli 1148	100%	0.0	523	99.61%	conjugal transfer protein TrbM
		Campylobacter coli 317/04	100%	0.0	523	99.61%	conjugal transfer protein TrbM
		Campylobacter sp. BCW_6462	100%	0.0	522	99.21%	conjugal transfer protein TrbM
		Campylobacter jejuni CVM 41934	100%	0.0	521	99.61%	conjugal transfer protein TrbM
45	WP_011117548.1	Campylobacter jejuni subsp. jejuni IA3902	100%	3.00E-162	456	100.00%	virulence protein
		Campylobacter jejuni subsp. jejuni 81-176	100%	9.00E-162	455	99.56%	virulence protein

		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
46	WP_011117549.1	Campylobacter jejuni subsp. jejuni 81-176	100%	0.0	727	100.00%	type IV secretion system protein VirB9
		Campylobacter jejuni subsp. jejuni IA3902	100%	0.0	725	99.72%	type IV secretion system protein VirB10
		Campylobacter coli	100%	0.0	725	99.72%	type IV secretion system protein VirB11
47	WP_011117575.1	Campylobacter jejuni subsp. jejuni IA3902	100%	0.0	592	100.00%	P-type conjugative transfer protein
		Campylobacter coli RM1875	100%	0.0	578	91.82%	P-type conjugative transfer protein
		Campylobacter jejuni X	100%	0.0	577	91.51%	P-type conjugative transfer protein
		Campylobacter coli 132-6	100%	0.0	576	91.19%	P-type conjugative transfer protein
		Campylobacter sp. B423b	100%	0.0	575	91.19%	P-type conjugative transfer protein
48	WP_011799393.1	*Campylobacter jejuni subsp. jejuni IA3902*	100%	0.0	536	100.00%	TrbM-like protein
		Campylobacter jejuni subsp. jejuni 81-176	100%	0.0	536	99.62%	TrbM-like protein
		Campylobacter coli	100%	0.0	535	99.62%	TrbM-like protein
49	WP_011117588.1	*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.

										Sub-cellular localization		Signal Peptide (Signal P)	Trans membrane helices prediction		
No	Protein IDs	No of Amino acid	MW	PI	Extinction coefficient	Instability Index	Classification	Alphabatic index	Grand average of Hydropathicity (GRAVY)	**CELLO**	**PSORT B**		**HMMTOP**	**TMHMM**	**SOSUI**
1	WP_002868767.1	400	43408	9.11	22350	26.24	Stable	87.52	-0.337	OuterMembrane	OuterMembrane	YES	No	No	Membrane, 1 TM helix
2	WP_002854524.1	140	16278.72	4.75	19285	29.26	Stable	93.43	-0.209	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
3	WP_009882162.1	241	26833.04	4.9	11920	35.93	Stable	43.9	-1.31	Extracellular	Unknown	No	No	No	Soluble
4	WP_010790856.1	136	15792.47	8.88	9190	28.8	Stable	77.5	-0.201	Cytoplasmic	Unknown	No	No	No	Soluble
5	WP_009882239.1	553	58834.74	4.79	33935	19.93	Stable	82.19	-0.41	Extracellular	OuterMembrane	YES	No	No	Soluble
6	WP_002854991.1	129	14888.85	5.31	11460	22.6	Stable	123.88	0.924	InnerMembrane	CytoplasmicMembrane	No	3 TM Helices	3 TM Helices	Membrane, 3 TM helix
7	WP_002855029.1	178	20660.76	4.93	34045	53.4	Unstable	101.4	-0.07	Cytoplasmic	Unknown	No	No	No	Soluble
8	WP_002868905.1	392	45037.15	9.63	52830	28.08	Stable	96.33	-0.377	OuterMembrane	Unknown	YES	No	No	Membrane, 1 TM helix
9	WP_002869356.1	360	41483.57	7.68	17880	31.74	Stable	98.06	-0.489	OuterMembrane	Cytoplasmic	No	1 TM Helices	1 TM Helices	Membrane, 1 TM helix
10	WP_002856929.1	238	27739.85	5.6	19160	40.81	Unstable	90.5	-0.789	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
11	WP_002869028.1	441	51113.83	6.5	41385	23.72	Stable	70.27	-0.744	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
12	WP_011812736.1	292	35243.21	9.33	29590	30.81	Stable	100.82	-0.395	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
13	WP_002868809.1	408	46384.58	5.39	54125	24.69	Stable	92.89	-0.256	OuterMembrane	Unknown	YES	No	No	Soluble
14	WP_002869368.1	150	16506.86	6.58	7615	27.75	Stable	64.33	-0.486	Cytoplasmic	OuterMembrane	No	No	No	Soluble
15	WP_0098825	448	52423.	9.2	66365	36.19	Stable	85.8	-0.439	OuterMemb	Unknown	No	No	No	Membra

	83.1		83	7						rane					ne, 1 TM helix
16	WP_002853389.1	272	31893.87	9.01	17545	42.12	Unstable	93.93	-0.638	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
17	WP_009882608.1	372	42214.61	4.8	20985	31.35	Stable	91.29	-0.388	Extracellular	OuterMembrane	No	No	No	Soluble
18	WP_002856369.1	138	15935.31	5.03	9970	38.87	Stable	111.67	-0.227	Cytoplasmic	Unknown	YES	1 TM Helices	No	Soluble
19	WP_079254190.1	57	7067.78	10.39	4470	32.55	Stable	114.56	0.053	Cytoplasmic	Unknown	YES	1 TM Helices	No	Membrane, 1 TM helix
20	WP_002856180.1	64	7495.58	4.49	125	23.56	Stable	100.31	-0.302	Cytoplasmic	Unknown	No	No	No	Soluble
21	WP_002831611.1	87	10262.55	6.07	18700	51.2	Unstable	63.79	-0.849	Cytoplasmic	Unknown	No	No	No	Soluble
22	WP_002790076.1	459	51016.24	5.27	13535	43.38	Unstable	107.52	-0.068	OuterMembrane	CytoplasmicMembrane	No	2 TM Helices	1 TM Helices	Membrane, 2 TM helix
23	WP_002853792.1	241	27299.29	5.12	19495	28.46	Stable	102.66	0.028	Cytoplasmic	Unknown	YES	No	No	Membrane, 1 TM helix
24	WP_002869072.1	316	37232.84	6.23	19160	24.49	Stable	83.35	-0.187	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
25	WP_002869097.1	452	52285.74	9.42	36010	35.76	Stable	93.89	-0.373	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
26	WP_002869326.1	464	52461.9	7.11	39310	27.14	Stable	93.73	-0.289	OuterMembrane	Unknown	No	1 TM Helices	No	Membrane, 1 TM helix
27	WP_002869139.1	779	90202.39	5.76	75360	43.31	Unstable	97.56	-0.315	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
28	WP_002869195.1	65	7142.96	7.98	2980	6.5	Stable	75.23	-0.806	Periplasmic	Unknown	No	No	No	Soluble
29	WP_002856630.1	788	93550.24	5.52	72020	36.88	Stable	93.68	-0.394	Cytoplasmic	Unknown	No	No	No	Soluble
30	WP_002855458.1	172	19575.46	5.31	4470	37.6	Stable	96.51	-0.472	Cytoplasmic	Cytoplasmic	No	1 TM Helices	No	Membrane, 1 TM helix
31	WP_002797496.1	142	16563.95	9.01	7450	40.15	Unstable	75.63	-0.985	Cytoplasmic	Cytoplasmic	No	1 TM Helices	No	Soluble
32	WP_024088174.1	237	27971.16	5.42	9190	49.19	Unstable	89.7	-0.393	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
33	WP_009883030.1	439	52369.05	5.66	48835	41.54	Unstable	103.87	-0.486	Cytoplasmic	Cytoplasmic	No	No	No	Soluble

34	WP_002824979.1	75	8773.25	6.27	5500	23.41	Stable	105.07	-0.167	Cytoplasmic	Unknown	No	No	No	Soluble
35	WP_002869225.1	288	32725.77	7.1	59610	30.26	Stable	113.75	0.844	InnerMembrane	CytoplasmicMembrane	No	8 TM Helices	8 TM Helices	Membrane, 8 TM helix
36	WP_002856602.1	138	15376.93	7.83	11460	17.73	Stable	98.91	-0.201	Periplasmic	Unknown	YES	No	No	Soluble
37	WP_002868888.1	328	39113.6	6.23	25370	26.75	Stable	95.43	-0.159	Cytoplasmic	Cytoplasmic	No	2 TM Helices	1 TM Helices	Membrane, 1 TM helix
38	WP_002868880.1	199	23051.39	8.95	24535	44.54	Unstable	101.01	-0.158	Extracellular	Unknown	No	2 TM Helices	No	Membrane, 1 TM helix
39	WP_009883121.1	166	18730.48	4.73	6085	23.14	Stable	99.94	-0.131	Cytoplasmic	Unknown	No	No	No	Soluble
40	WP_002860117.1	286	32643.66	5.19	37360	28.98	Stable	103.78	-0.095	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
41	WP_002779704.1	88	10623.3	7.53	17795	57.44	Unstable	81.93	-0.601	Cytoplasmic	Unknown	No	No	No	Soluble
42	WP_011187233.1	408	47059.95	9.1	48025	28.47	Stable	79.39	-0.746	Cytoplasmic	Cytoplasmic	No	No	No	Soluble
43	WP_011187235.1	597	69031.68	7.94	51020	29.19	Stable	91.57	-0.431	OuterMembrane	Cytoplasmic	No	No	No	Soluble
44	WP_002809111.1	254	29365.92	8.88	39475	41.3	Unstable	79.84	-0.534	Periplasmic	Cytoplasmic	YES	No	No	Soluble
45	WP_011117548.1	225	25915.78	7.78	18910	24.92	Stable	93.16	-0.18	OuterMembrane	Unknown	No	2 TM Helices	1 TM Helices	Membrane, 1 TM helix
46	WP_011117549.1	356	40874.42	8.83	32320	34.91	Stable	80.06	-0.613	OuterMembrane	Unknown	YES	No	No	Soluble
47	WP_011117575.1	292	33294.59	5.67	19495	34.99	Stable	69.18	-0.692	Periplasmic	Unknown	YES	1 TM Helices	1 TM Helices	Soluble
48	WP_011799393.1	260	30029.93	8.54	60445	30.85	Stable	70.15	-0.637	OuterMembrane	Unknown	YES	1 TM Helices	No	Soluble
49	WP_011117588.1	134	15715.47	9.82	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_002854991.1	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_002869356.1	Ketol-acid reductoisomerase (EC-1.1.1.86) (340 aa)	0.882
10	WP_002856929.1	No result	NA
11	WP_002869028.1	ulfate-binding protein precursor (348 aa)	0.884
12	WP_011812736.1	Sensor protein ZraS (EC-2.7.13.3) (339 aa)	0.859
13	WP_002868809.1	NADP-dependent 3-hydroxy acid dehydrogenase YdfG (EC-1.1.1.-) (249 aa)	0.888
14	WP_002869368.1	hermonuclease precursor (EC-3.1.31.1) (175 aa)	0.533
15	WP_009882583.1	ATP-dependent RecD-like DNA helicase (EC-3.6.4.12) (447 aa)	0.674
16	WP_002853389.1	Membrane protein insertase YidC	0.889
17	WP_009882608.1	ABC transporter glutamine-binding protein GlnH precursor (279 aa)	0.643
18	WP_002856369.1	Rhomboid protease AarA (EC-3.4.21.105) (157 aa)	0.614
19	WP_079254190.1	No result	NA
20	WP_002856180.1	Cadmium, cobalt and zinc/H(+)-K(+) antiporter (316 aa)	0.875
21	WP_002831611.1	Uncharacterized protein (64 aa)	0.681
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_002869072.1	Murein DD-endopeptidase MepM (EC-3.4.24.-) (273 aa)	0.859
25	WP_002869097.1	3-oxoacyl-[acyl-carrier-protein] synthase 3 (EC-2.3.1.180) (353 aa)	0.86
26	WP_002869326.1	1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC-1.1.1.267)	0.862
27	WP_002869139.1	Pyruvate-flavodoxin oxidoreductase (EC-1.2.7.-)	0.979
28	WP_002869195.1	Uncharacterized protein (144 aa)	0.712
29	WP_002856630.1	ATP-dependent helicase/nuclease subunit A (EC-3.1.-.-) (921 aa)	0.991
30	WP_002855458.1	Adenylosuccinate synthetase (EC-6.3.4.4)	0.883
31	WP_002797496.1	Adenylosuccinate synthetase	0.911
32	WP_024088174.1	Formate dehydrogenase	0.893
33	WP_009883030.1	Uncharacterized protein (282 aa)	0.859
34	WP_002824979.1	NADH dehydrogenase subunit C	0.987
35	WP_002869225.1	Anaerobic dimethyl sulfoxide reductase chain A (774 aa)	0.956
36	WP_002856602.1	Uncharacterized protein (247 aa)	0.754
37	WP_002868888.1	Uncharacterized protein (183 aa)	0.88

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

Table S9. 3D structural information of HPs from C. Jejuni

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	2c5uA	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

			CHEMOTAXIS PROTEIN						
23	WP_002853792.1	No result							
24	WP_002869072.1	1zkdA	DUF185	73.10%	*Rhodopseudomonas palustris*	X-RAY DIFFRACTION	2.1 Å	0.258	0.223
25	WP_002869097.1	No results							
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		0.404
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	1zbmA	hypothetical protein AF1704	79.92%	*Archaeoglobus fulgidus*	X-RAY DIFFRACTION	2.3 Å	0.265	0.212
41	WP_002779704.1	2cm5A	RABPHILIN-3A	72.22%	*Rattus norvegicus*	X-RAY DIFFRACTION	1.28 Å	0.194	
42	WP_011187233.1	2au3A	DNA primase	58.81%	*Aquifex aeolicus*	X-RAY DIFFRACTION	2 Å	0.238	0.203
43	WP_011187235.1	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			

					typhimurium				
47	WP_011117575.1	2ch7A	METHYL-ACCEPTING CHEMOTAXIS PROTEIN	81.99%	*Thermotoga maritima*	X-RAY DIFFRACTION	2.5 Å	0.297	0.259
48	WP_011799393.1	1zsoB	hypothetical protein	73.97%	*Plasmodium falciparum*	X-RAY DIFFRACTION	2.17 Å	0.233	0.183
49	WP_011117588.1	1ne8A	conserved hypothetical protein YDCE	56.56%	*Bacillus subtilis*	X-RAY DIFFRACTION	2.1 Å	0.210	0.159