



www.bioinformatics.net
Volume 18(3)

Research Article

Received October 21, 2021; Revised March 11, 2022; Accepted March 31, 2022, Published March 31, 2022

DOI: 10.6026/97320630018188

Declaration on Publication Ethics:

The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at <https://publicationethics.org/>. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Declaration on official E-mail:

The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

License statement:

This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Comments from readers:

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Edited by P Kanguane

Citation: Sharma *et al.* Bioinformatics 18(3): 188-195 (2022)

Functional assignment to hypothetical proteins in *Orientia tsutsugamushi* strain Ikeda

Dixit Sharma^{1,*}, Sunil Kumar¹, Ankita Sharma², Rakesh Kumar¹, Ranjit Kumar¹, Mahesh Kulharia² & Manish Kumar³

¹Department of Animal Sciences, School of Life Sciences, Central University of Himachal Pradesh, District Kangra, Himachal Pradesh, India - 176206; ²Centre for Computational Biology and Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, District Kangra, Himachal Pradesh, India - 176206; ³Department of Chemistry and Chemical Sciences, School of Physical and Material Science, Central University of Himachal Pradesh, District Kangra, Himachal Pradesh, India - 176206; *Corresponding author

Author contacts:

Dixit Sharma - E-mail: sharmadixit17@hpcu.ac.in; sharmadixit17@gmail.com

Sunil Kumar - E-mail: sunilibes@hpcu.ac.in

Ankita Sharma - E-mail: sharma889ankita@gmail.com

Rakesh Kumar - E-mail: drthakurcuhp@hpcu.ac.in

Ranjit Kumar - E-mail: ranjitcuhp@hpcu.ac.in

Mahesh Kulharia - E-mail: kulharia@hpcu.ac.in

Manish Kumar - E-mail: manishssu26@hpcu.ac.in

Abstract:

Orientia tsutsugamushi (*O. tsutsugamushi*) is an intracellular bacterial pathogen which causes zoonosis scrub typhus in humans. Genome of *O. Tsutsugamushi* strain Ikeda contains 214 hypothetical proteins (HPs) which is nearly 20% of the total proteins. Domain and family based functional analysis of HPs results in the annotation of 44 hypothetical proteins. The annotated HPs were classified in to five main classes namely, gene expression and regulation, transport, metabolism, cell signaling and proteolysis. Thus, computational analysis of HPs helps to understand their putative roles in various biological and cellular processes, including pathogenesis for further consideration as potential therapeutic targets.

Keywords: *Orientia tsutsugamushi*, hypothetical proteins, therapeutic targets

Background:

Orientia tsutsugamushi (*O. tsutsugamushi*) is an obligate intracellular pathogen responsible for infectious illness scrub typhus in humans. The zoonotic disease scrub typhus is a vector borne disease transmitted to human through the bite of *Leptotrombidium* mite larvae. The disease is predominant in Asia-Pacific region but the cases from Middle East and South America showed the wide geographical distribution of pathogen[1,2]. Scrub typhus in South East Asia is most important emerging infectious zoonotic disease and most common rickettsial disease in India. The outbreak of the disease was reported in different provinces in India including Assam, Himachal Pradesh, Jammu and Kashmir, Bihar, Rajasthan, Tamilnadu and may more [3]. The infection of *O. tsutsugamushi* can also lead to multiorgan failure and other disease like neurological complications, Parkinson's disease, acute kidney injury and acute respiratory distress syndrome [4-6]. Due to high mortality and non-specific representation of disease the effectual control measure are needed. Earlier attempts were made to prioritize metabolic pathway proteins and metal-binding proteins of *O. tsutsugamushi* as potential therapeutic targets [7,8]. The experimental techniques are very useful but expensive and time consuming whereas computational techniques evolved as most promising and reliable approaches for dissecting the genome or proteome sequence datasets[9,10]. Earlier, computational methods were efficiently used for deciphering the whole proteome data of different bacterial pathogens [11,12]. Therefore, it is of interest to document data on the functional assignment to hypothetical proteins (HPs) in *Orientia tsutsugamushi* strain Ikeda for further consideration in drug development.

Methodology:**Hypothetical proteins dataset retrieval:**

The complete proteome dataset of *O. tsutsugamushi* strain Ikeda was downloaded from Ref Seq server. Further, the HPs were extracted by searching the term "hypothetical proteins" in the whole proteins dataset.

Functional prediction:

Functional conserved signature sequences or domains of HPs were predicted by Inter Pro Scan [13] and Pfam [14]. Inter Pro Scan allows protein sequence to search against Inter Pro's signatures *i.e.*, predicted models that define families, domains and sites of the proteins. Pfam is the collection of curated protein families with clearly described organization of domains.

Metabolic pathway analysis:

Blast KOALA [15] at Kyoto Encyclopedia of Genes and Genomes (KEGG) was used for the analysis of metabolic pathways of HPs. KEGG is integrated database for biological interpretation of genomic and other high throughput sequence data.

Sub-cellular localization prediction:

The subcellular localization of HPs were investigated by SOSUIGramN [16]. SOSUIGramN predict subcellular localization on the basis of physicochemical parameters of N- and C-terminal signal sequence and total sequence. The overall accuracy of the software is 92.9%.

Identification of potential therapeutic targets among HPs:

To extract non-homologous proteins of *O. tsutsugamushi*, BLASTp [17] search of HPs was performed against host (human; taxid: 9906) proteins at NCBI with expect value (E-value) 10^{-5} . Bacterial toxins are described as most potent targets for drug discovery process. Essentiality of HPs was checked with the help of Database of Essential Genes (DEG) version 10 [18]. The DEG harbors essential genes and proteins, regulatory sequences, noncoding RNAs, promoters and replication origins. The BLASTp search of the HPs was performed against DEG database with cut-off parameter of 0.00001 E-value. The selected hypothetical proteins having homologous hits with the DEG proteins database were considered as essential. Expasy's Prot Param [19] was used for the physical and chemical parameters analysis of HPs from *O. tsutsugamushi*. The identification of physicochemical properties of proteins provides insight into their biochemical nature and serves as the initial step for the studies like drug discovery. Further, the prediction of bacterial toxins among HPs was performed by BTXpred [20]. BTXpred is the online server to predict bacterial toxins with overall accuracy of 96.07%.

Results and Discussion:

The vector borne zoonotic disease scrub typhus caused by an intracellular pathogen *O. tsutsugamushi* is most remarkable neglected tropical disease having ever-widening impact, worldwide. Studies have even shown that in addition to scrub typhus *O. tsutsugamushi* can lead to various other complications [4,5]. Hypothetical proteins are the proteins with known amino acid sequences but with unknown functions. Earlier studies showed the involvement of HPs in different biological processes including drug development [5]. The whole proteome analysis of *O. Tsutsugamushi* strain Ikeda showed that about 20% of the total proteins are hypothetical. We have functionally annotated and classified 44 HPs based on conserved functional domain and family. The 44

functionally annotated HPs were classified in to five broad classes, *i.e.*, gene expression and regulation, cell signaling, metabolism, transport and proteolysis. The detailed description of the study is mentioned below.

The *O. tsutsugamushi* strain Ikeda proteome was publically available at RefSeq with genome reference number NC_010793.1. The size of the genome is 2.01 Mb with 30.5% GC content. The proteome of *O. tsutsugamushi* contains 1227 proteins out of which 241 were HPs (**Figure 1**). The 241 HPs were extracted from the whole proteome and selected for further analysis. The HPs are encoded by open reading frames but they do not have any known function. In many pathogenic bacteria, HPs play crucial roles in pathogens survival and advancement of related infectious diseases [23].

The hypothetical proteins were annotated on the basis of conserved domain and family. We have identified the function of 44 HPs and classified in to five broad classes, *i.e.*, gene expression and regulation (25), transport (9), metabolism (5), cell signaling (4) and proteolysis (1) (**Figure 2**). The main domains present in the HPs were DnaA N-terminal domain, Outer membrane enzyme PagP, Integrase and Group II intron (**Figure 3**). Genome sequencing of *O. tsutsugamushi* showed that it contains immense number of repetitive sequences[24]. In our present study we find the existence of integrase, ankyrinrepeat and DnaB-like proteins which belong to

the category of repetitive sequences. The *in vivo* gene expression of *O. tsutsugamushi* was studied earlier and identified genes found to involved in DNA replication, protein translation, post translational modification, signaling, transport and metabolism [25]. The study strengthens our results as we have also observed the presence of proteins in class of gene expression and regulation, cell signaling, metabolism and transport.

The category of gene expression and regulation showed the presence of DnaA N-terminal domain, integrase, group II intron, transposase DDE domain, DnaB-like, Leucine-rich repeat domain, Phage gp-6-like, ribosomal protein S2, RNA polymerase, ribonuclease reductase-like and ribonuclease H superfamily (**Figure 3, Table 2**). N-terminal domain is crucial for the activity of bacterial DnaA at the origin of replication. It is involved in the coordination of chromosomal replication with cell cycle, cellular processes and oxidative stress [26]. Replicative DnaB-like helicases are required for unwinding of DNA strand and important in the movement of replication fork direction. The C-terminal region of DnaB helicases is vital for proper folding and contains motifs necessary for helicases activity [27]. Furthermore integrase mediates specific cut and paste recombination between two DNA recognition sequences [28]. The versatile and compact group-II introns of bacteria are known to create genetic diversity from a population of intron invaded mRNAs by circulation and trans-splicing to produce new splicing products with novel functions [29].

Table 1: Function of 44 hypothetical proteins from *O. tsutsugamushi* on domain and family analysis. The 44 HPs were broadly classified in to five broad classes, *i.e.*, gene expression and regulation, transport, metabolism, cell signaling and proteolysis.

S.No.	Protein ID	Subcellular localization	Functional domain	Broad class
1	WP_012460960.1	Cytoplasmic	DnaA N-terminal domain	Gene expression and regulation
2	WP_012461007.1	Unknown	DnaA N-terminal domain	Gene expression and regulation
3	WP_012461012.1	Outer Membrane	DnaA N-terminal domain	Gene expression and regulation
4	WP_012461214.1	Outer Membrane	DnaA N-terminal domain	Gene expression and regulation
5	WP_012461800.1	Unknown	DnaA N-terminal domain	Gene expression and regulation
6	WP_080503960.1	Cytoplasmic	DnaA N-terminal domain	Gene expression and regulation
7	WP_080503973.1	Cytoplasmic	DnaA N-terminal domain	Gene expression and regulation
8	WP_080503986.1	Cytoplasmic	DnaA N-terminal domain	Gene expression and regulation
9	WP_148141126.1	Cytoplasmic	DnaA N-terminal domain	Gene expression and regulation
10	WP_155367750.1	Unknown	Integrase	Gene expression and regulation
11	WP_158296904.1	Unknown	Integrase	Gene expression and regulation
12	WP_041621600.1	Cytoplasmic	Integrase	Gene expression and regulation
13	WP_041621651.1	Cytoplasmic	Integrase	Gene expression and regulation
14	WP_080503988.1	Unknown	Group II intron	Gene expression and regulation
15	WP_158296914.1	Cytoplasmic	Group II intron	Gene expression and regulation
16	WP_158296923.1	Unknown	Group II intron	Gene expression and regulation
17	WP_148141124.1	Unknown	DnaB-like	Gene expression and regulation
18	WP_012461582.1	Outer Membrane	Leucine-rich repeat domain	Gene expression and regulation
19	WP_012461489.1	Cytoplasmic	Phage gp6-like	Gene expression and regulation
20	WP_080503912.1	Unknown	Ribonuclease H superfamily	Gene expression and regulation
21	WP_012461954.1	Cytoplasmic	Ribonucleotide reductase-like	Gene expression and regulation
22	WP_041621580.1	Inner Membrane	Ribosomal protein S2	Gene expression and regulation
23	WP_041621591.1	Cytoplasmic	RNA polymerase subunit	Gene expression and regulation
24	WP_080503914.1	Unknown	Transposase DDE domain	Gene expression and regulation
25	WP_108883745.1	Unknown	Transposase DDE domain	Gene expression and regulation
26	WP_012461016.1	Outer Membrane	Outer membrane enzyme PagP	Transport
27	WP_012461958.1	Outer Membrane	Outer membrane enzyme PagP	Transport
28	WP_012462005.1	Outer Membrane	Outer membrane enzyme PagP	Transport
29	WP_012462151.1	Outer Membrane	Outer membrane enzyme PagP	Transport
30	WP_050731353.1	Outer Membrane	Outer membrane enzyme PagP	Transport
31	WP_012461356.1	Unknown	TolA/TonB	Transport
32	WP_012461335.1	Extracellular	Armadillo-type fold	Transport
33	WP_041621668.1	Inner Membrane	MrpF_PhaF	Transport
34	WP_012462081.1	Cytoplasmic	P-loop NTPase	Transport

35	WP_012461347.1	Cytoplasmic	Alpha/Beta hydrolase fold	Metabolism
36	WP_041621613.1	Extracellular	Citrate synthase	Metabolism
37	WP_012462252.1	Inner Membrane	Glycoside hydrolase family 18	Metabolism
38	WP_012461669.1	Extracellular	Lipase_3	Metabolism
39	WP_012461162.1	Unknown	Uroporphyrinogen III synthase superfamily	Metabolism
40	WP_012461997.1	Extracellular	Ankyrin repeat	Cell Signaling
41	WP_012462020.1	Cytoplasmic	HD-domain	Cell Signaling
42	WP_041621676.1	Inner Membrane	Methyl-accepting chemotaxis protein	Cell Signaling
43	WP_012461311.1	Unknown	Bacterial SH3 domain	Cell Signaling
44	WP_012462337.1	Extracellular	Ulp1 protease	Proteolysis

Table 2: List of 44 hypothetical proteins with sequence length and molecular weight. All proteins are non-homologous to host proteome. Out of these 37 were predicted as bacterial toxins and 5 were essential.

S.No.	Protein ID	Sequence length (a.a.)	Molecular weight (kDa)	Homology with host	Essential	Bacterial toxin
1	WP_012460960.1	519	60.701	X	X	Yes
2	WP_012461007.1	165	19.316	X	X	Yes
3	WP_012461012.1	571	67.070	X	X	Yes
4	WP_012461016.1	182	20.406	X	X	Yes
5	WP_012461162.1	227	25.802	X	X	Yes
6	WP_012461214.1	571	67.029	X	X	Yes
7	WP_012461311.1	165	19.543	X	Yes	Yes
8	WP_012461335.1	1126	124.602	X	X	Yes
9	WP_012461347.1	286	32.535	X	X	Yes
10	WP_012461356.1	409	45.249	X	Yes	X
11	WP_012461489.1	181	20.962	X	X	Yes
12	WP_012461582.1	440	50.170	X	X	Yes
13	WP_012461669.1	397	43.928	X	X	Yes
14	WP_012461800.1	160	18.528	X	X	Yes
15	WP_012461954.1	115	13.686	X	X	Yes
16	WP_012461958.1	167	18.774	X	X	Yes
17	WP_012461997.1	521	58.792	X	X	Yes
18	WP_012462005.1	262	29.092	X	X	X
19	WP_012462020.1	363	42.948	X	X	Yes
20	WP_012462081.1	216	25.237	X	Yes	X
21	WP_012462151.1	182	20.242	X	X	Yes
22	WP_012462252.1	372	43.320	X	Yes	Yes
23	WP_012462337.1	1368	153.611	X	X	Yes
24	WP_041621580.1	258	28.956	X	X	Yes
25	WP_041621591.1	80	8.985	X	X	X
26	WP_041621600.1	65	7.871	X	X	Yes
27	WP_041621613.1	101	11.482	X	Yes	Yes
28	WP_041621651.1	105	12.482	X	X	X
29	WP_041621668.1	84	9.827	X	X	Yes
30	WP_041621676.1	256	27.638	X	X	Yes
31	WP_050731353.1	237	26.055	X	X	Yes
32	WP_080503912.1	48	5.643	X	X	Yes
33	WP_080503914.1	47	5.730	X	X	Yes
34	WP_080503960.1	249	28.896	X	X	Yes
35	WP_080503973.1	165	19.395	X	X	Yes
36	WP_080503986.1	449	52.806	X	X	Yes
37	WP_080503988.1	52	6.566	X	X	Yes
38	WP_108883745.1	38	4.763	X	X	Yes
39	WP_148141124.1	40	4.686	X	X	Yes
40	WP_148141126.1	367	42.753	X	X	Yes
41	WP_155367750.1	57	6.709	X	X	X
42	WP_158296904.1	58	6.692	X	X	X
43	WP_158296914.1	62	7.411	X	X	Yes
44	WP_158296923.1	45	5.224	X	X	Yes

In the class of cell signaling we found the presence of ankyrin repeat, methyl-accepting chemotaxis protein (MCP), bacterial SH3 domain and HD-domain (**Figure 3**). The ankyrin repeat motif is nearly 33 residues long and is the most common protein-protein interaction motif in nature. The interaction of ankyrin repeat with target proteins is mediated by forming helix-turn-helix structure within protein [30]. Ankyrin repeats are functionally diverse and in addition to *O. tsutsugamushi*, also present in other intracellular pathogenic bacteria (*Coxiellaburnetii*, *Rickettsiaspp.*, and *Wolbachiapipientis*) [31]. Various intracellular bacterial pathogens

translocate ankyrin repeat containing proteins by different secretion systems into host cells, where they manipulate various host processes [30,32]. The chemotaxis MCP protein act as receptor for intracellular and environmental signals and regulates various cellular signaling processes involved in cell survival, pathogenicity and biodegradation [33].

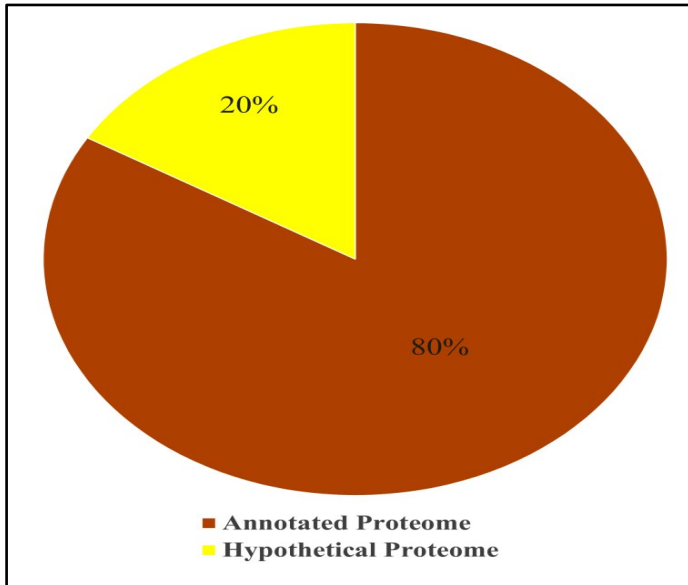


Figure 1: The representation of hypothetical proteome in the total proteome of *O. tsutsugamushi*. The 20% of the proteome is hypothetical and 80% is already annotated.

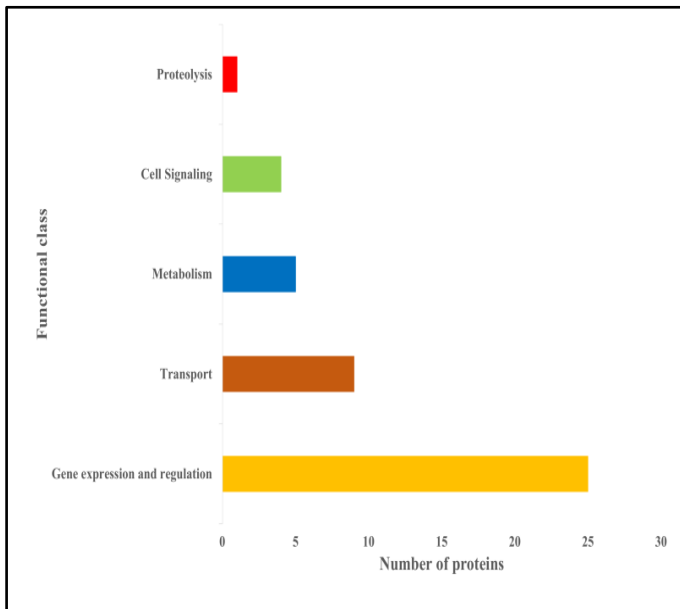


Figure 2: Functional classification of *O. tsutsugamushi* hypothetical proteins. The 44 annotated hypothetical proteins were classified into five broad classes, i.e., gene expression and regulation, transport, metabolism, cell signaling and proteolysis. Most of the proteins are present in the class of gene expression and regulation and transport.

The metabolism class showed the presence of citrate synthase, lipase, glycoside hydrolase family 18, alpha/beta hydrolase fold and uroporphyrinogen III synthase (Figure 3, Table 1). The enzyme of kerbs cycle or central energy metabolism, i.e., citrate synthase is a moonlighting enzyme which act an essential check point

transcriptional regulator to control bacterial cell cycle[34]. Lipase is a type of hydrolases which hydrolyzes the ester bonds in monoacylglycerols and have crucial function in fat metabolism [35]. The role of enzyme GH18 is known earlier in hydrolyzing glycosidic bond present between two or more carbohydrates [36].

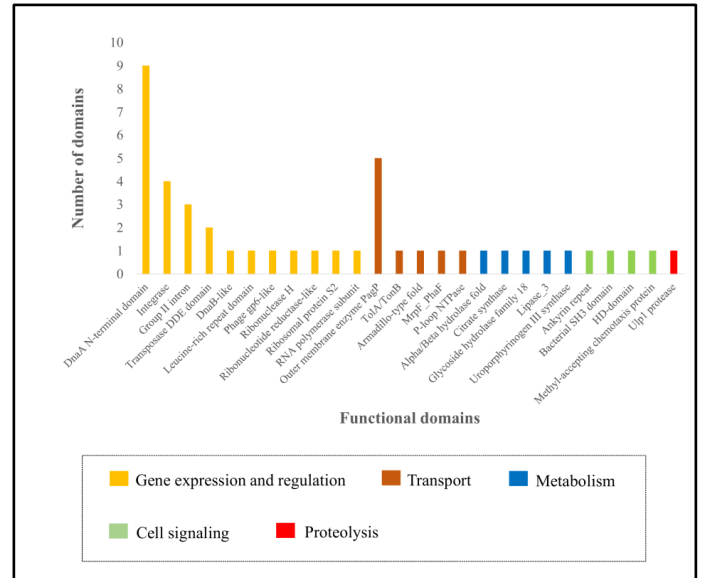


Figure 3: The graphical representation of functional domains of hypothetical proteins and their distribution in different broad functional classes. The X-axis represents functional domains and Y-axis represents number of functional domains. The color of each column indicates its broad class.

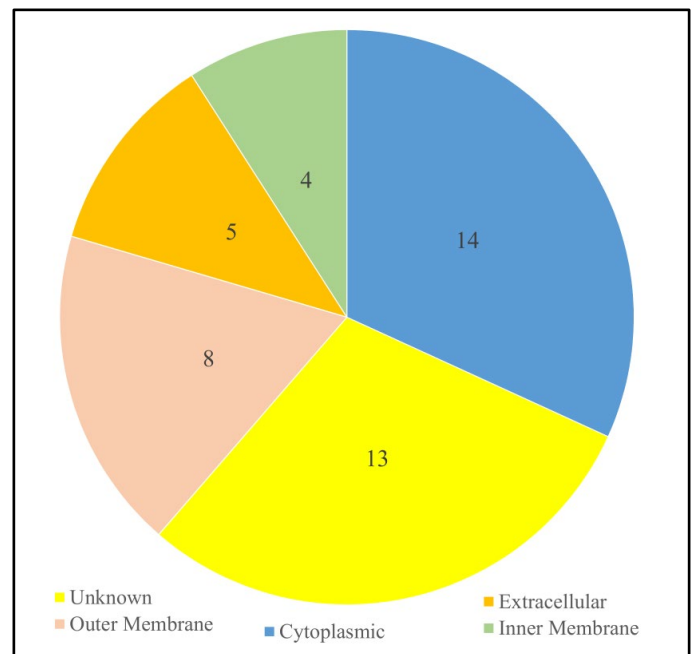


Figure 4: Subcellular localization of *O. tsutsugamushi* hypothetical proteins.

The transport of category consists of outer membrane enzyme PagP, TolA, MrpF_PhaF, P-loop NTPase and Armadillo-type fold (**Figure 3, Table 1**). The outer membrane enzyme PagP in gram-negative bacteria is responsible for transferring palmitate chain from phospholipid to glucosamine unit of lipid A, cationic antimicrobial peptides resistance, signal transduction and in evading host immune defense [37,38]. TolA protein interacts with Pal protein to form an inner membrane and outer membrane connection which ultimately involved in antibiotic resistance [39]. MrpF_PhaF is an integral membrane protein which act as a part of potassium efflux pump that involved in regulation of pH [40]. The P-loop NTPase is one of the most abundant monophyletic protein domain present in the proteomes of prokaryotes. The role of P-loop NTPases have noticed earlier in numerous processes including transcription, replication, membrane transport, intracellular trafficking and activation of different metabolites [41,42]. Ulp1 protease (ubiquitin-like protein-specific proteases) is the single protein found in the class of proteolysis (**Figure 3**). Ulp1 is the major proteases which cleave small ubiquitin-related modifiers (SUMO) and play essential role in progression of cell cycle [43,44].

The metabolic pathway analysis of 44 HPs was carried out at KEGG with the help of BlastKOALA. Two HPs of *O. tsutsugamushi*, i.e., WP_04161591.1 (*rpoZ*, DNA-directed RNA polymerase subunit omega) and WP_041621613.1 (*gltA*, citrate synthase) found to be involved in different metabolic pathways. Protein WP_04161591.1 with KEGG orthology number K03060 is involved in RNA polymerase pathway with KEGG pathway ID map03020. The role of *rpoZ* gene has been noticed earlier in various physiological processes such as formation of biofilm, motility and resistance to antibiotics [45]. The protein WP_041621613.1 with KEGG orthology number K01647 is involved in eight pathways, namely citrate cycle (KEGG pathway ID: map00020), glyoxylate and dicarboxylate metabolism (KEGG pathway ID: map00630), metabolic pathways (KEGG pathway ID: map0100), biosynthesis of secondary metabolites (KEGG pathway ID: map01110), microbial metabolism in diverse environments (KEGG pathway ID: map01120), carbon metabolism (KEGG pathway ID: map01200), 2-oxocarboxylic acid metabolism (KEGG pathway ID: map01210) and biosynthesis of amino acids (KEGG pathway ID: map01230). The *gltA* is required by pathogenic bacteria to establish the infection in different organs. Earlier, the citrate synthase is known to conferred metabolic flexibility to pathogen *Klebsiella pneumonia* which impact site specific fitness during infection [46]. Previous studies reported that the proteins involved in the various metabolic pathways of pathogen may act as probable therapeutic targets and are crucial for pathogens survival [7,47,48]. Therefore, HPs involved in metabolic pathways can serve as the probable drug targets and help in the drug discovery process.

The distribution of functional HPs in the subcellular locations of *O. tsutsugamushi* was predicted by SOSUIGramN. The subcellular localization of 44 HPs showed that 14 were present in cytoplasm, 4 in inner membrane, 8 in outer membrane, 5 in extra cellular space and 13 HPs have unknown location. **Figure 4** shows the subcellular location of 44 HPs. The precise subcellular localization is an

important authentication used for inferring function of the protein. It can also help in designing the lab experiment to study particular protein of interest and may help in the identification of antimicrobial targets, factors contributing to host invasion and potential vaccine candidates [49].

The identification of non-human homologous HPs will enhance the specificity of the designed drug and targeting of host homologous proteins may cause cytotoxicity and cross-reactivity [7]. Hence, we prioritized potential therapeutic targets first by identifying the non-human homologous HPs in *O. tsutsugamushi*. All the 44 functionally annotated HPs were non-human homologous i.e. were present in *O. tsutsugamushi* and absent in host *Homo sapiens* (**Table 2**). Essential proteins are imperative for the survival of the organism. The DEG was used to predict the essential proteins in the set of functionally annotated HPs. We have identified 17 HPs which may be essential for the survival of the *O. tsutsugamushi*. Five HPs (WP_012461311.1, WP_012461356.1, WP_012462081.1, WP_012462252.1 and WP_041621613.1) were found essential which may involve in numerous cellular functions and have crucial role in the survival of the pathogen [47].

The physicochemical properties of the HPs were predicted further which are crucial for their structure, activity and biological function. The molecular weight of 44 HPs ranges from 4.68 kDa (WP_148141124.1) to 153.61 kDa (WP_012462337.1). The proteins molecular weight should be ≤ 110 kDa to act better therapeutic targets as earlier studies suggested that small sized proteins are more soluble and can be easily purified in wet lab experiments [50,51]. Two HPs (WP_012461335.1 and WP_012462337.1) have molecular weight more than 110 kDa. A total of 29 HPs have theoretical pI more than 7 and 15 HPs had theoretical pI less than 7. The predicted pI of HPs can be useful to prepare the buffers for protein purification using isoelectric focusing. The aliphatic index of the HPs ranges from 73 (WP_148141124.1) to 160.12 (WP_041621668.1). The high aliphatic index indicates that these proteins can be stable at wide range of temperature [52,53]. The GRAVY of the HPs were calculated and it ranges from -0.834 (WP_080503986.1) to 1.323 (WP_041621668.1). The low value of GRAVY is the positive indicator for the good interaction of water molecules with proteins. The host cell functions were modulated by bacterial toxins (virulence factor) in order to favor microbial infection [54]. There are different virulence factors which can be cytosolic, secretory and membrane associated. These virulence factors are important constituents of bacteria as they help to undergo morphological and physiological shifts, aid in adherence to host cell or to kill host cells and to resist the host immune response [55][56]. We have predicted that 37 non-human homologous HPs were putative virulent factors that may contribute to pathogenesis of *O. tsutsugamushi* (**Table 2**). Furthermore, the functional annotation performed in earlier step also strengthen the finding that HPs proteins can act as putative therapeutic targets and play important roles in drug development.

Conclusion and future perspectives:

O. tsutsugamushi is an intracellular pathogen of one of the neglected disease scrub typhus and therefore its genome is less studied by the scientists. More than one quarter of the proteins in *O. tsutsugamushi* strain Ikeda have unknown function. We report 44 annotated proteins among 241 HPs using systematic Bioinformatics tools. Most of the annotated HPs involved in gene expression and regulation and cell signaling. The categories of metabolism, transport and proteolysis have shared few proteins. These proteins found to be localized in cytoplasm, inner membrane, outer membrane and extra cellular space. Moreover, all the annotated 44 proteins found non-homologous to human and 37 proteins among these 44 proteins have shown virulence activity implying therapeutic nature for further consideration in drug discovery against the pathogen.

Author contribution:

Dixit Sharma, Sunil Kumar and Mahesh Kulharia: Conceived the idea and designed the experiment; Dixit Sharma and Ankita Sharma: Annotation and classification of hypothetical proteins; Rakesh Kumar, and Manish Kumar: Data analysis and curation; Dixit Sharma: Writing original-draft preparation; Ankita Sharma, Ranjit Kumar, Manish Kumar and Mahesh Kulharia: Checked and edited the manuscript for important intellectual content; All authors reviewed the manuscript and gave final approval for publication.

Funding:

The research is supported by Department of Health Research, Government of India in the form of Young Scientist Fellowship to Dixit Sharma (Grant Number: YSS/2020/000116/PRCYSS).

Acknowledgement:

Dixit Sharma acknowledges Department of Health Research, Government of India for providing Young Scientist Fellowship. Authors duly acknowledge Central University of Himachal Pradesh for providing research facilities.

Conflict of interest:

The authors declare no conflict of interest.

References:

- [1] Weitzel T *et al.* *N Engl J Med* 2016 **375**:954. [PMID:27602667]
- [2] Sharma D *et al.* *Microb Pathog* 2021 **158**: 105103. [PMID:34298125]
- [3] Sarma N *Indian J Dermatol* 2017 **62**:451. [PMID:28979005]
- [4] Mahajan SK *et al.* *Emerg Infect Dis* 2010 **16**:1641. [PMID:20875305]
- [5] Premaratna R *et al.* *BMC Res Notes* 2015 **8**:438. [PMID:26369561]
- [6] Xu G *et al.* *PLoS Negl Trop Dis* 2017 **11**:e0006062 [PMID:29099844]
- [7] Sharma D *et al.* *J Mol Recognit* 2018 **32**:e2766. [PMID:30343521]
- [8] Sharma D *et al.* *Front Genet* 2019 **10**:797. [PMID:31608099]
- [9] Sharma A *et al.* *Bio Metals* 2017 **30**:127. [PMID:28105572]
- [10] Sharma A *et al.* *Front Microbiol* 2018 **9**:1838 [PMID:30186242]
- [11] Sharma A *et al.* *Physiol Mol Plant Pathol* 2019 **106**:187. <https://doi.org/10.1016/j.pmpp.2019.02.005>
- [12] Sharma A *et al.* *R Soc Open Sci* 2019 **6**:190369. [PMID:31598288]
- [13] Jones P *et al.* *Bioinformatics* 2014 **30**:1236 [PMID:24451626]
- [14] Finn RD *et al.* *Nucleic Acids Res* 2015 **44**:D279. [PMID:26673716]
- [15] Kanehisa M *et al.* *J Mol Biol* 2016 **428**:726. [PMID:26585406]
- [16] Imai K *et al.* *Bioinformatics* 2008 **2**:417 [PMID:18795116]
- [17] Altschul SF *et al.* *J Mol Biol* 1990 **215**:403 [PMID:2231712]
- [18] Luo H *et al.* *Nucleic Acids Res* 2014 **42**:D574. [PMID:24243843]
- [19] Gasteiger E *et al.* *Nucleic Acids Res* 2003 **31**:3784. [PMID:12824418]
- [20] Saha S & Raghava GPS. *In Silico Biol* 2007 **7**:405. [PMID:18391233]
- [21] Sharma D & Bisht D. *Front Microbiol* 2017 **8**:465. [PMID:28377758]
- [22] Singh B *et al.* *Futur Sci* 2017 **3**:66. [PMID:28344832]
- [23] Desler C *et al.* *BMC Bioinformatics* 2009 **10**:289. [PMID:19754976]
- [24] Nakayama K *et al.* *DNA Res* 2008 **15**:185. [PMID:18508905]
- [25] Cho B *et al.* *Proteomics* 2010 **10**:1699. [PMID:20186754]
- [26] Zawilak-Pawlik A *et al.* *Genes(Basel)* 2017 **8**:136. [PMID:28489024]
- [27] Nitharwal RG *et al.* *Nucleic Acids Res* 2007 **35**:2861. [PMID:17430964]
- [28] Groth AC & Calos MP. *J Mol Biol* 2004 **335**:667. [PMID:14687564]
- [29] LaRoche-Johnston F *et al.* *PLoS Genet* 2018 **14**:e1007792. [PMID:30462638]
- [30] Al-Khodori S *et al.* *Trends Microbiol* 2010 **18**:132. [PMID:19962898]
- [31] Voth DE *et al.* *J Bacteriol* 2009 **191**:4232. [PMID:19411324]
- [32] VieBrock L *et al.* *Front Cell Infect Microbiol* 2015 **4**:186. [PMID:25692099]
- [33] Ud-Din AIMS & Roujeinikova A. *Cell Mol Life Sci* 2017 **74**:3293. [PMID:28409190]
- [34] Berge M *et al.* *Elife* 2020 **9**:e52272 [PMID:32149608]
- [35] Derewenda ZS. *Adv Protein Chem Elsevier* 1994 **45**:1. [PMID:8154368]
- [36] Monge EC *et al.* *J Biol Chem* 2018 **293**:3849. [PMID:29367339]
- [37] Bishop RE, *Mol Microbiol* 2005 **57**:900. [PMID:16091000]
- [38] Iyer BR & Mahalakshmi R, *Biochemistry* 2016 **55**:4960. [PMID:27525547]
- [39] Egan AGF, *Mol Microbiol* 2018 **107**:676. [PMID:29315884]
- [40] Yang C *et al.* *Protein Expr Purif* 2012 **85**:60. [PMID:22750396]
- [41] Leipe DD *et al.* *J Mol Biol* 2003 **333**:781. [PMID:14568537]
- [42] Aravind L *et al.* *Genome Biol* 2004 **5**:R30. [PMID:15128444]
- [43] Li SJ & Hochstrasser M, *Nature* 1999 **398**:246 [PMID:10094048]
- [44] Gareau JR & Lima CD, *Nat Rev Mol Cell Biol* 2010 **11**:861 [PMID:21102611]

- [45] Bhardwaj N *et al. Genes to Cells* 2018 **23**:357. [PMID:29542212]
- [46] Vornhagen J *et al. PLoS Pathog* 2019 **15**:e1008010. [PMID:31449551]
- [47] Sharma A & Pan A. *Eur J Med Chem* 2012 **57**:185. [PMID:23059547]
- [48] Sharma OP & Kumar MS. *Sci Rep* 2016 **6**:19842 [PMID:26806463]
- [49] Rey S *et al. BMC Genomics* 2015 **6**:162. [PMID:16288665]
- [50] Duffield M *et al. MolBiosyst* 2010 **6**:2482. [PMID:20949199]
- [51] Parvege MM *et al. Drug Target Insights* 2014 **8**:DTI-S19728. [PMID:25574133]
- [52] Ikai A, *J Biochem* 1980 **88**:1895. [PMID:7462208]
- [53] Guo X *et al. PLoS One* 2016 **11**:e0164810. [PMID:27741324]
- [54] do Vale A *et al. Front Microbiol* 2016 **7**:42. [PMID:26870008]
- [55] Sharma AK *et al. Indian J Microbiol* 2017 **57**:1. [PMID:28148975]
- [56] Aktories K & Barbieri JT. *Nat Rev Microbiol* 2005 **3**:397. [PMID:15821726]

