



## OPEN ACCESS

## EDITED BY

Lalima K. Madan,  
Medical University of South Carolina,  
United States

## REVIEWED BY

Kamil Steczkiewicz,  
Polish Academy of Sciences, Poland  
Stefan Simm,  
Universitätsmedizin Greifswald,  
Germany  
Sebastien Pichoff,  
University of Kansas Medical Center,  
United States

## \*CORRESPONDENCE

Andrew M. Lynn,  
✉ andrew@jnu.ac.in

## SPECIALTY SECTION

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 03 August 2022

ACCEPTED 05 December 2022

PUBLISHED 04 January 2023

## CITATION

Kaur H and Lynn AM (2023), Mapping the  
FtsQBL divisome components in  
bacterial NTD pathogens as potential  
drug targets.  
*Front. Genet.* 13:1010870.  
doi: 10.3389/fgene.2022.1010870

## COPYRIGHT

© 2023 Kaur and Lynn. This is an open-  
access article distributed under the  
terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which does  
not comply with these terms.

# Mapping the FtsQBL divisome components in bacterial NTD pathogens as potential drug targets

Harbinder Kaur and Andrew M. Lynn\*

School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

Cytokinesis is an essential process in bacterial cell division, and it involves more than 25 essential/non-essential cell division proteins that form a protein complex known as a divisome. Central to the divisome are the proteins FtsB and FtsL binding to FtsQ to form a complex FtsQBL, which helps link the early proteins with late proteins. The FtsQBL complex is highly conserved as a component across bacteria. Pathogens like *Vibrio cholerae*, *Mycobacterium ulcerans*, *Mycobacterium leprae*, and *Chlamydia trachomatis* are the causative agents of the bacterial Neglected Tropical Diseases Cholera, Buruli ulcer, Leprosy, and Trachoma, respectively, some of which seemingly lack known homologs for some of the FtsQBL complex proteins. In the absence of experimental characterization, either due to insufficient resources or the massive increase in novel sequences generated from genomics, functional annotation is traditionally inferred by sequence similarity to a known homolog. With the advent of accurate protein structure prediction methods, features both at the fold level and at the protein interaction level can be used to identify orthologs that cannot be unambiguously identified using sequence similarity methods. Using the FtsQBL complex proteins as a case study, we report potential remote homologs using Profile Hidden Markov models and structures predicted using AlphaFold. Predicted ortholog structures show conformational similarity with corresponding *E. coli* proteins irrespective of their level of sequence similarity. Alphafold multimer was used to characterize remote homologs as FtsB or FtsL, when they were not sufficiently distinguishable at both the sequence or structure level, as their interactions with FtsQ and FtsW play a crucial role in their function. The structures were then analyzed to identify functionally critical regions of the proteins consistent with their homologs and delineate regions potentially useful for inhibitor discovery.

## KEYWORDS

neglected tropical disease, divisome, FtsQBL, remote homologs, profile HMMs, AlphaFold

## 1 Introduction

Cytokinesis is an essential step of cell division, and errors in this process may lead to cell death. To carry out this process accurately, bacteria employ a highly conserved and complex machinery known as the divisome. The divisome is a protein complex made up of more than 25 proteins (Du and Lutkenhaus, 2017), some of which are essential for cytokinesis. These essential proteins include early proteins (FtsA, FtsZ, ZipA), which form a proto-ring/Z-ring, and late proteins (FtsK, FtsQ, FtsL, FtsB, FtsW, FtsI, FtsN), which are recruited to the proto-ring (Aarsman et al., 2005; den Blaauwen et al., 2017; Söderström and Daley, 2017). Central to the divisome are the proteins FtsB and FtsL binding to FtsQ to form a complex FtsQBL, which helps link the early proteins with late proteins (Choi et al., 2018). Among the late proteins, FtsW and FtsI are critical components of peptidoglycan synthesis (Mercer and Weiss, 2002). FtsQ is essential and interacts with many other divisome components, which makes it an excellent target for cell division inhibitors (Buddelmeijer and Beckwith, 2004; Kureisaite-Ciziene et al., 2018). Both FtsQBL and FtsWI complexes are highly conserved across bacteria. Recruitment of the FtsWI complex depends on the cytoplasmic domain of FtsL (Gonzalez et al., 2010; Park et al., 2020) which is a component of the FtsQBL complex. Components of the divisome are excellent drug targets due to their essentiality. Highly homologous proteins would serve as targets for the design of broad host-range antibiotics, while remote homologs with more sequence divergence may serve as specific targets. The FtsW binds downstream to penicillin-binding proteins involved with peptidoglycan synthesis. These are well-studied and used antibiotic targets.

Neglected tropical diseases (NTDs) are prevalent in low-income economic regions of Asia, Africa, and the Americas. They are caused by diverse pathogens such as bacteria, viruses, protozoa, and helminths (Daumerie et al., 2010). Bacterial NTDs like Cholera, Buruli Ulcerans, Leprosy, and Trachoma are caused by pathogens *Vibrio cholerae*, *Mycobacterium ulcerans*, *Mycobacterium leprae*, and *Chlamydia trachomatis*, respectively. Cholera is a primeval disease that causes severe diarrhea due to the consumption of contaminated water and unhygienic living conditions (Faruque et al., 1998). The Buruli ulcer mainly affects the skin but can also affect bones, resulting in permanent disability (WHO, 2022a). The mode of transmission is not yet known, and treatment is costly, though it is believed to spread through personal contact. Leprosy is another chronic disease caused by the bacteria *Mycobacterium leprae* (WHO, 2022b). Many multidrug therapies are available, and the widespread global presence of the disease was reduced by 90% in three decades from 1985 (Kealey, 2010). But the treatment is not easily available to the very poor, and victims may continue to suffer from social stigma, disability, and disfigurement. Trachoma is an infection generally occurring in the eyelids caused by *Chlamydia trachomatis* (WHO, 2022c). It transmits through

the discharge released from the eye of an infected person. Reinfection can occur and can result in visual impairment or complete blindness (Kealey, 2010). In comparison to other diseases, very few drugs have been discovered for NTDs recently (WHO, 2022c).

Interestingly, while *V. cholerae* has a self-sufficient genome contributing to its ability to survive in aquatic reservoirs outside of the host, *Mycobacterium* and *Chlamydia* have reduced genomes corresponding to their obligate intracellular parasitic nature (Stephens et al., 1998; Cole et al., 2001; Stinear et al., 2007). *Mycobacterium sp.* and *Chlamydia* seemingly lack many components of divisome which are shown to be essential in a model organism like *E. coli*. In *Mycobacterium sp.*, some early proteins of divisome that helps in the correct assembly of FtsZ appear to be missing (Ouellette et al., 2020) until the *sepF* gene was identified (Gola et al., 2015), which interacts with mycobacterial FtsZ protein and alteration of which caused a division defect in *Mycobacterium smegmatis* that led to filamentous cells. *Chlamydia* has eliminated many unnecessary genes and processes and kept only the genes crucial for the bacterium to evolve into intracellular parasites. Surprisingly, some of the genes lost include a number of essential Fts cell division genes, including FtsZ (Stephens et al., 1998). In the past, (Ouellette et al., 2012) discovered and provided evidence that *Chlamydia* uses proteins that determine rod shape for cell division. They also proposed that MreB replace FtsZ in the division process. Following that, (Kemege et al., 2015) displayed MreB localization information at the cell division site for *Chlamydia*. Due to the deletion or replacement of these essential proteins, these pathogens have evolved a non-canonical divisome, with known homologs to canonical divisome proteins not detected using standard sequence similarity. The presence of FtsQ, FtsB, and FtsL homologs and their assembly into complexes in gram-positive bacteria like *Bacillus subtilis* and *Streptococcus pneumoniae* suggests that the FtsQBL subcomplex is evolutionarily conserved (Daniel et al., 2006). Beckwith's group (2010) (Gonzalez et al., 2010) conducted a bioinformatics evolutionary analysis based on 16s rRNA sequences in 400 genomes and found that homologs of FtsQ, FtsB, and FtsL *E. coli* proteins were present in most of the organisms using a combination of HMMs with PFAM profiles and synteny. Results from this study also indicate the presence of all three proteins in the *Mycobacterium* and the presence of FtsL in *Chlamydia*.

These four bacterial NTDs represent genome diversity within bacterial phyla, including genome reduction. Ortholog mapping is an active area of research, with a growing number of methods that are chosen based on a user's need for accuracy, speed, available computational resources, size of the application datasets or requirement for integration within a pipeline (Nichio et al., 2017). The Quest for Orthologs benchmark service provides a single framework to evaluate multiple publicly available methods (Nevers et al., 2022). Most

methods are based on sequence similarity coupled with a higher-order graph or tree-based clustering for inference. There is a need for a standard method to propagate annotation from well-studied model organisms to identify orthologs irrespective of their sequence divergence. In this paper, a protocol is described to characterize distant remote homologs of *E. coli* proteins FtsQ (ecFtsQ), FtsB (ecFtsB), and FtsL (ecFtsL) in bacterial NTDs. The initial step for protein function prediction is often a sequence similarity search against the sequences of the known function. BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1997) is widely used for sequence similarity search against non-redundant databases or customized databases. However, sequence-sequence comparison methods are unable to find the homologs in the target organisms that have very low sequence similarity. To search for the remote homologs of the FtsQBL complex of cell division proteins in *Mycobacterium* and *Chlamydia*, sequence-profile (Finn et al., 2011) and profile-profile Hidden Markov Models (HMMs) (Söding, 2005) was used.

The increased sensitivity of HMM-HMM methods comes at the cost of specificity-as remote homologs can share a common fold, but not necessarily common molecular function. These proteins can be differentiated either by sequence similarity to a homolog functionally characterized from an evolutionarily closer organism or from the protein structure. To more accurately predict the protein's function and differentiate between proteins with a common fold, analysis of the protein's structure can be added to the functional annotation pipeline. AlphaFold (Jumper et al., 2021) is an Artificial Intelligence (AI) based program for predicting protein structures with high accuracy and speed developed by Google's DeepMind. Predicted structures of an ortholog will show conformational similarity with corresponding *E. coli* proteins despite having low sequence similarity. The fine function of proteins, which share almost the same domain organization leading to high structural similarity, can also be mapped from their structural characteristics. To distinguish between homologs that share a common structural motif or fold, their differential interactions with partner proteins in a complex could play an important role; FtsB and FtsL provide such a condition, being small bitopic amphipathic helices, each allowing sequence diversity without affecting their function, making them difficult to predict with accuracy using sequence similarity methods. The periplasmic region of both of them binds to the periplasmic region of FtsQ while the cytoplasmic region of FtsL binds to FtsW. To understand the interactions between the FtsQBL complex and the interactions between FtsL and FtsW, AlphaFold multimer (Evans et al., 2022) was used. AlphaFold multimer predicts the structure for multi-chain protein complexes while maintaining intrachain accuracy of the structure. Structures of many components of the divisome from multiple bacteria are known, however, the structure of the entire complex remains unknown with few structures of interacting multimers. Crystal structures of FtsQ (2VH1) (van den Ent et al., 2008), the coiled-coil segment of FtsB (residues 28-

63) (4IFF) (Lapointe et al., 2013), and FtsQB periplasmic complex (6H9N) (Kureisaite-Ciziene et al., 2018) from *E. coli* are present in the PDB database. As multimer models were used to predict function, the *E. coli* multimer structure was modeled with AlphaFold v2 and compared with known experimentally determined interaction to validate the method. Although FtsB and FtsL are unannotated in Uniprot for *M. ulcerans* and *M. leprae*, a previous study by Wu et al. (2018) have characterized FtsB and FtsL homologs for *Mycobacterium smegmatis*. These proteins were characterized with the help of HMM models and *in vivo* studies. Ouellette et al. (2015) characterized homologs of FtsQ and FtsL in *Chlamydia* with the help of bacterial Y2H assay, to date, these entries are still listed as hypothetical and uncharacterized in Uniprot along with FtsB homolog. The results from these studies were used to validate our approach.

## 2 Materials and methods

### 2.1 Data retrieval

The proteome sequences of all the four organisms, i.e., *Vibrio cholerae* (RefSeq ID: GCF\_000016245.1), *Chlamydia trachomatis* (RefSeq ID: GCF\_000008725.1), *Mycobacterium ulcerans* (RefSeq ID: GCF\_000013925.1), and *Mycobacterium leprae* (RefSeq ID: GCF\_000195855.1) were downloaded from the NCBI dataset. The complete genome filter in the NCBI dataset was applied and earliest completed genome was used for each organism in our analysis. Annotation and protein sequences of FtsQ, FtsB, FtsL, and FtsW homologs in *E. coli* were retrieved from the UniProt database (Bateman et al., 2021) with Uniprot ID P06136, P0A6S5, P0AEN4, and P0ABG4, respectively.

### 2.2 Identification of homolog and remote homologs

The remote homologs of the FtsQ, FtsB, FtsL, and FtsW from *E. coli* were identified in all the four organisms with the sequential use of BLAST (Altschul et al., 1997), HMMER (Finn et al., 2011), and HHSearch (Söding, 2005). The programs were installed locally using instructions from the distribution website. For BLAST, the *E. coli* sequences were used as query proteins, while the proteome files of the individual organisms were separately formatted as BLAST databases. For profile HMMs, ortholog profiles of query proteins from *E. coli* i. e. COG1589 for FtsQ, COG2919 for FtsB and COG3116 for FtsL, were retrieved from EggNOG5 (Huerta-Cepas et al., 2019) and searched against target organisms' protein sequences using "hmmsearch" from the HMMER version 3.2 package (Finn et al., 2011). Profile-profile mapping was carried out with the help of HH-suite3 package v3.0.3 (Steinegger et al., 2019). Multiple sequence alignment (MSA) profiles of query proteins and proteome files were generated

with “HHblits” (Remmert et al., 2011) while performing two iterations with the Uniprot20 (version 2016), a clustered version of the UniProt database, which works well for ortholog detection. It may be noted that the default database currently distributed with HH-suite is the UniRef-30 and earlier Uniclust-30 (Mirdita et al., 2017), which performs well for the programs common use of detection of remote homologs at the fold level but not for our purpose of ortholog mapping. “HHmake” was used to convert the MSAs into profiles of the hhm format. A database was similarly built from the proteome file of target organisms *Chlamydia*, *Mycobacterium ulcerans*, and *Mycobacterium leprae*.

## 2.3 Structure prediction

AlphaFold v2.0 (Jumper et al., 2021) was implemented for structure prediction of the potential homologs for proteins FtsQ, FtsB, FtsL, and FtsW. AlphaFold was installed locally along with all the required genetic (sequence) databases using instructions from DeepMind’s GitHub repository (Jumper et al., 2021). A reduced version of all the databases (BFD, MGnify, PDB, Uniclust30, Uniprot, and UniRef 90) was used with the database preset option. The reduced\_db preset has been optimized for speed and low hardware requirements. For individual proteins, AlphaFold monomer was used to predict the structure of remote homologs of FtsQ, FtsL, FtsB, and FtsW. To understand the multimeric interfaces to form the FtsQBL complex and recruitment of FtsW by FtsL, AlphaFold multimer (Evans et al., 2022) was used. AlphaFold Multimer requires a multifasta file as input. For generating both monomer and multimer models max\_template\_date = 2020-05-14 was used. Both models generated five models and ranked them on the basis of the plddt score.

## 2.4 Structure analysis

Structural alignment was performed between predicted structures of potential homologs with corresponding *E. coli* homologs using the STAMP alignment tool (Russell and Barton, 1992) from MultiSeq extension (Roberts et al., 2006) in VMD (Visual Molecular Dynamics) (Humphrey et al., 1996). Interactions between the multimeric protein complexes were calculated using the web server PDBSum (Laskowski et al., 2018). Images of superimposed structures of both monomers and multimers were generated using VMD (Humphrey et al., 1996).

## 2.5 Multiple sequence alignment and phylogenetic analysis

Orthologous sequences for each predicted homolog of FtsQ, FtsB, and FtsL proteins in all four organisms were extracted from

the EggNOG v5 (Huerta-Cepas et al., 2019) database on the basis of the COG (clusters of orthologous groups) to which they belong. FtsQ and FtsB homologs mapped to COG1589 and COG2919, and two COGs were found for FtsL protein COG3116 and COG4839. The sequences corresponding to these COGs were downloaded, and the individual sequences were added for predicted homologs of A0PTJ5, O84041 and O84273 that did not map to any of the standard FtsB and FtsL COGs. A multiple alignment was built using the MAFFT v7.5 (Katoh and Standley, 2013) tool with default parameters. FastTree v2.1.11 (Price et al., 2010) was used to generate a phylogenetic tree from the multiple alignments using default parameters for protein sequences. A sequence logo was generated from MSAs with the help of Weblogo 3 (Crooks et al., 2004).

## 3 Results

### 3.1 Sequence similarity methods can predict potential candidate homologs in NTD bacterias with varying degrees of confidence

Homologs of ecFtsQ, ecFtsB, ecFtsL, and ecFtsW proteins were easily mapped in *V. cholerae* based on sequence similarity with BLAST. Single candidate proteins with low e-values (i.e., <0.001) (Table 1) were mapped as homologs to the *E. coli* query proteins. Probable peptidoglycan polymerase FtsW is highly conserved across bacteria and was also easily mapped using BLAST in all four organisms. For *Mycobacterium* and *Chlamydia*, a methodical study was used to find a remote homolog for FtsQ, FtsB, and FtsL as BLAST was not able to detect any significant hits for these proteins. The HMM-based method HMMER identified candidate homologs for FtsQ, FtsB and FtsL in *M. ulcerans* and *M. leprae* with significant e-value (i.e., <0.001) but was not able to provide any significant hit for *C. trachomatis*. Predicted FtsQ homologs in *M. ulcerans* A0PTI5 (muFtsQ) and *M. leprae* Q9CCE5 (mlFtsQ) (Table 1) are already annotated and present in Uniprot (Bateman et al., 2021) under the unreviewed annotation section. Both the homologs have slightly bigger sizes than the *E. coli* protein. The sequence-profile method identified two potential homologs for FtsB and one potential homolog for FtsL for each *Mycobacterium* sp. (Table 1). To identify the other proteins, a more sensitive method, profile-profile comparison, was used with a probability cutoff of 0.95 (HH-suite User Guide; Söding, 2005). This method successfully identified candidate proteins for remote homologs of ecFtsQ, ecFtsB, and ecFtsL in *Mycobacterium* sp and *Chlamydia*. The homolog identified as *Chlamydia* FtsQ (ctFtsQ) (Table 1) was annotated as a hypothetical protein in UniProtKB (O84769). Its sequence length (268 residues) is slightly shorter than the *E. coli* homolog.

Profile-profile methods identified two potential homologs each for both ecFtsB and ecFtsL in *Mycobacterium* and

**TABLE 1** Results for FtsQ, FtsB, FtsL, and FtsW remote homologs mapped on target genomes. For BLAST and HMMER, only e-values are listed against the proteins. For HHsearch, both e-values and probability scores are listed.

Organism	Method	FtsQ	FtsL		FtsB		FtsW
<i>V. cholerae</i>		<b>Q9KPG9</b>	<b>Q9KPG0</b>		<b>Q9KUJ3</b>		<b>Q9KPG6</b>
	<b>Blast</b>	3e-56	4e-21		1e-21		2e-164
<i>M. ulcerans</i>		<b>A0PTI5</b>	<b>A0PW54*</b>	<b>A0PTJ5*</b>	<b>A0PW54*</b>	<b>A0PTJ5*</b>	<b>A0PTI8</b>
	<b>Blast</b>	-	-	-	-	-	1e-43
	<b>HMMER</b>	1.6e-40	2.5e-05	-	3e-26	1.5e-17	7.5e-50
	<b>HHsearch</b>	1.2E-29, 99.9	1.8E-15, 99.2	4.2E-11, 98.5	8E-18, 99.4	7.5E-13, 98.8	7.1E-58, 100
<i>M. leprae</i>		<b>Q9CCE5</b>	<b>Q9CD41*</b>	<b>Q7AQC6*</b>	<b>Q9CD41*</b>	<b>Q7AQC6*</b>	<b>Q7AQC4</b>
	<b>Blast</b>	-	-	-	-	-	3e-42
	<b>HMMER</b>	3.3e-40	3e-05	-	1.5e-23	4.7e-25	2.7e-49
	<b>HHsearch</b>	4.3E-30, 99.9	2E-15, 99.2	8.3E-12, 98.6	8E-18, 99.4	6.4E-13, 98.8	1.8E-63, 100
<i>C. trachomatis</i>		<b>O84769*</b>	<b>O84041*</b>	<b>O84273*</b>	<b>O84041*</b>	<b>O84273*</b>	<b>O84765</b>
	<b>Blast</b>	-	-	-	-	-	3e-36
	<b>HMMER</b>	-	-	-	-	-	5.6e-38
	<b>HHsearch</b>	2E-29, 99.9	7.4E-18, 99.4	2.1E-16, 99.3	9.3E-19, 99.5	4.4E-17, 99.4	1.1E-63, 100

The asterisk (\*) sign indicates the proteins were unannotated in Uniprot.

The Bold text represents the Uniprot ids of candidate homologs and name of the Method used. This is purely stylistic and was used to distinguish them from numerical values.

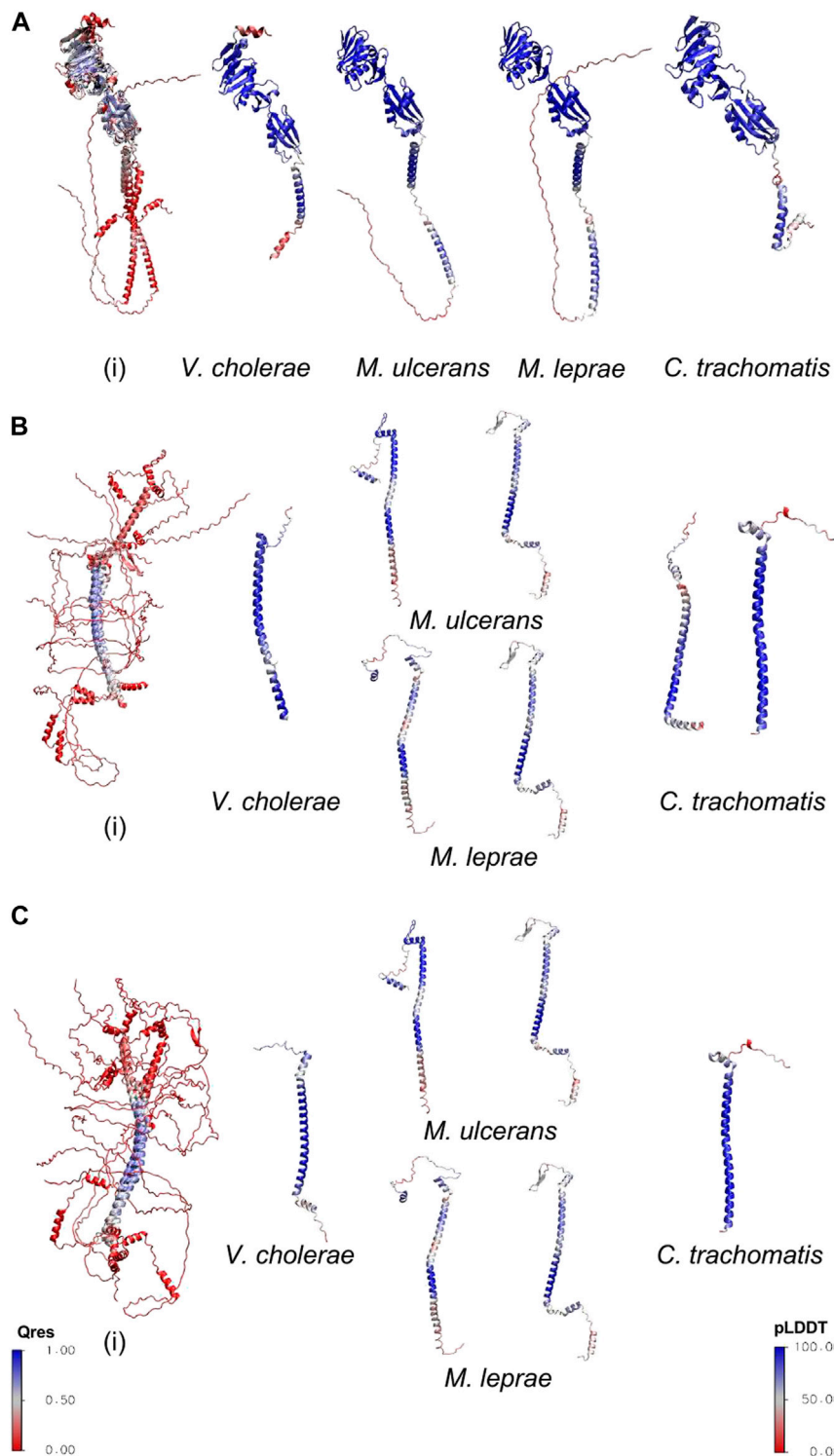
*Chlamydia*. These candidate homologs identified by both methods are same for FtsB and FtsL and hence these proteins are not significantly distinguishable using remote homology methods in both *Mycobacterium* sp. and *Chlamydia*. Although the size of these homologs is almost identical in *Chlamydia* compared to corresponding *E. coli* proteins, in *Mycobacterium* sps. it is much larger. It is reasonable to obtain ambiguous hits between these two proteins using profile-profile comparison algorithms, given that FtsB and FtsL are both small bitopic, amphipathic proteins with coiled-coil domains (Condon et al., 2018) and show low sequence conservation. Protein structure prediction was carried out to characterize these candidate remote homologs based on structural characteristics to solve the ambiguity between FtsB and FtsL.

### 3.2 Structural similarity is an additional validation of function inferred from remote sequence similarity

Candidate remote homologs of ecFtsQ, ecFtsB, and ecFtsL were identified using Sequence-Profile and Profile-Profile comparison methods. Although their sequences are not very close to those of *E. coli* proteins, when structural modeling was performed using AlphaFold, similar structures and conserved domains were detected.

FtsQ is a bitopic membrane protein composed of 276 amino acids and possesses the POTRA domain in the periplasmic

region, which is crucial for its recruitment of binding partners (van den Ent et al., 2008). AlphaFold provides a confidence metric pLDDT which measures the accuracy of the predicted models. The models with pLDDT  $\geq$  90 represent prediction with high accuracy and between 70 and 90 represent a good backbone prediction (Jumper et al., 2021). The disordered regions were excluded from the calculation of the pLDDT score. The pLDDT score for *V. cholerae* FtsQ (vcFtsQ), muFtsQ, mIFtsQ and ctFtsQ are 89.9, 90, 84.7, and 88.30 respectively. The homologs for this protein in *Mycobacterium* sps. and *Chlamydia* are composed of almost the same number of amino acids: 317 and 268, respectively. Predicted FtsQ homolog for *M. ulcerans* shows a domain organization similar to *E. coli*, possessing a transmembrane domain (104-124 amino acids), and POTRA domain (128-196), but in contrast to *E. coli*, it also contains an extended cytoplasmic disordered region (1-61). Similarly, the ctFtsQ homolog contains a large periplasmic domain in addition to the periplasmic and transmembrane domain in the same conformation as of *E. coli* protein. In comparison to ecFtsQ, both homologs contain a similar conformation with the same number of helices and  $\beta$ -sheets in the periplasmic domain except for the last C-terminus helix. Structural alignment was carried out with STAMP (Russell and Barton, 1992) a tool integrated into the Mutiseq extension of VMD (Humphrey et al., 1996) to measure the overall structure conservation of these remote homologs. These models of homologs were superimposed with *E. coli* structure which was



**FIGURE 1**

(A) The predicted structures of homologs superimposed with the experimental structure of *E. coli* FtsQ (PDB ID 2vh1). (B) The predicted structures of candidate homologs superimposed with FtsB homolog of *E. coli* (AlphaFold Model). (C) The predicted structures of remote homologs superimposed with FtsL homolog of *E. coli* (AlphaFold Model). The superimposed structures have been colored on the basis of structural conservation measures (Q-value) with blue being conserved and red variability between their structures. The individual structures of *V. cholerae*, *M. ulcerans*, *M. leprae*, and *C. trachomatis* are colored on the basis of AlphaFold confidence score.

experimentally determined (PDB ID: 2vh1) (van den Ent et al., 2008) as shown in Figure 1A. The superimposed models of vcFtsQ and ctFtsQ have Qres values 0.7 and 0.5 respectively. The muFtsQ and mlFtsQ have 0.4 Qres value due to N-terminus disordered region. The periplasmic domain of predicted FtsQ homologs from *Chlamydia* and *Mycobacterium* exhibits significant structural conservation, which is crucial for forming a complex with FtsB and FtsL. However, there is some variation in the size of the last helix towards the C-terminus, which is followed by two  $\beta$ -sheets and residues that form the interacting surface with FtsB. This helix is truncated due to a deleted segment in *Mycobacterium* sp. and *Chlamydia* compared to the proteobacteria ecFtsQ and vcFtsQ.

FtsB and FtsL are both small bitopic proteins with a size of 121 and 103 amino acids in *E. coli*, respectively. Both proteins contain a coiled-coil structural motif. In all three organisms, HHsearch results show ambiguity in identifying FtsB and FtsL proteins. AlphaFold was used to predict structures of these detected homologs for FtsB and FtsL in all three organisms. The pLDDT measure for structural model of *V. cholerae* (vcFtsB) and *V. cholerae* (vcFtsL) are 85.5 and 84.3 respectively. In *Mycobacterium* sp. the structural models have large disordered regions and the confidence score was calculated excluding these regions. The confidence score for the candidate homologs for FtsB and FtsL is 81.8 and 81.4 in *M. ulcerans*; 81.03 and 72.8 for *M. leprae*. In *Chlamydia* the confidence score is 77.4 and 93.1 for the predicted homologs. Detected homologs of FtsB and FtsL in *M. leprae* and *M. ulcerans* are twice the sizes of ecFtsB and ecFtsL (Table 1) having large disordered regions in both N-termini (1-62 residues and 1-93 residues) as well as C-termini (205-227 residues and 350-377 residues) regions, respectively. Both the proteins possess a coiled-coil domain (118 and 145, 153-180 residues) and the transmembrane helical region (90-112 and 122-142). Similarly, *M. ulcerans* homologs also contain disordered N-termini (1-52 and 1-95 residues) and C-termini (199-233 and 213-328 residues). Coiled-coil (113-133 and 153-180 residues) and transmembrane regions (85-107 and 118-142 residues) are also present in both of the detected hits. In *Chlamydia*, the sequence length of predicted homologs for FtsB and FtsL is very close to *E. coli*. Still, due to similar domain organization, it is challenging to differentiate between FtsB and FtsL. Both the bitopic coiled-coil models contain the helical transmembrane (20-38 residues) region. To distinguish between FtsB and FtsL homologs for *Mycobacterium* sp. and *Chlamydia*, models of these candidate homologs were aligned with ecFtsB and ecFtsL (AlphaFold models) (Figures 1B, C) for all three organisms as the experimentally determined structure for the *E. coli* proteins has not been determined. The candidate homologs showed structural alignment with FtsB (Figure 1B). Similarly, FtsL showed structural conservation with all the candidate homologs in the *Mycobacterium*, but in *Chlamydia*, it

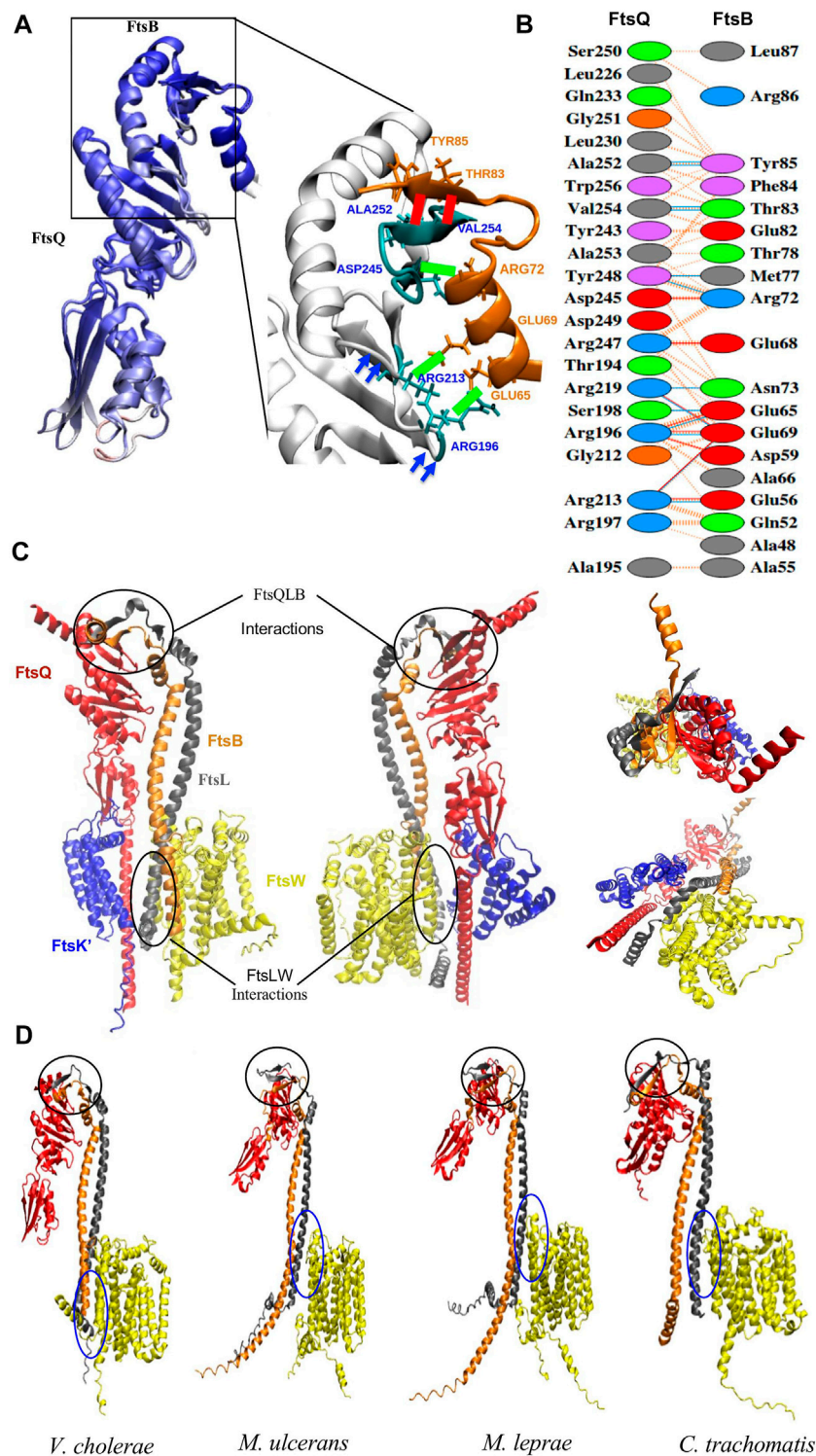
showed structural alignment with only one protein (O84273) (Figure 1C). The overall Qres score for *Mycobacterium* sp. is low due to large disordered region on N and C termini but the coiled-coil domains show structural conservation. Based on these structural characteristics and alignment, it is difficult to distinguish between the remote homologs for FtsB and FtsL. FtsB and FtsL form a subcomplex and then bind to the periplasmic domain of FtsQ, and the cytoplasmic region of FtsL interacts with FtsW for the recruitment of FtsW to the septum site (Gonzalez et al., 2010). Interactions of FtsQB and FtsLW were studied from the modeled multimer complex of FtsK', FtsQ, FtsL, FtsB, and FtsW to distinguish between the candidate homologs of FtsB and FtsL.

The structural model of FtsW, a polytopic membrane protein with ten transmembrane segments, is very similar across all the predicted homologs in all the organisms mentioned above. It shows minor variation in the first helical region (N-terminus) in *Vibrio* and *Mycobacterium*, and this helical region is absent in *Chlamydia*.

### 3.3 The *E. coli* FtsKQBLW complex serves as a reference for intermolecular interactions in bacteria

The predicted remote homologs for FtsB and FtsL are ambiguous in *Mycobacterium* sp. and *Chlamydia*. To differentiate these remote homologs as FtsB or FtsL, their selective interactions with FtsQ and FtsW respectively are pivotal (Gonzalez et al., 2010; Du and Lutkenhaus, 2017). The structure of the FtsQBL complex and the recruitment of FtsW by cytoplasmic FtsL is not fully understood. In *E. coli* FtsQBL protein complex occurs in stable conformations of trimeric (1:1:1) complex and hexameric (2:2:2) complex (Villanelo et al., 2011). So far, only the periplasmic FtsQ and FtsB subcomplex, have their bound structures determined experimentally (PDB ID 6H9N) (Kureisaite-Ciziene et al., 2018). AlphaFold multimer was used to predict the structure (Figure 2) of *E. coli* FtsK', FtsQ, FtsL, FtsB, and FtsW to understand the interactions between these proteins. Three of these proteins (FtsQ, FtsB, and FtsL) each have a single transmembrane helix near their N-termini, while FtsW is a polytopic protein and contains 10 transmembrane segments and many loops within the cytoplasm that could interact with the cytoplasmic domain of FtsL (Pastoret et al., 2004). FtsQBL forms a complex independent of FtsK or FtsW. The transmembrane region of FtsK' (up to 180 residues) was used to anchor the N-terminus of the FtsQ in the membrane, and to prevent this domain from interfering with the FtsBL interactions with FtsW.

Most of the critical interactions for the binding of FtsB and FtsL occur in the periplasmic region of FtsQ. In the model, the C-termini of both FtsB and FtsL form a strand-like structure only when bound to FtsQ. The AlphaFold modeled FtsQB was



**FIGURE 2**

(A) Superimposition of the ecFtsQB PDB structure (6H9N) with the AlphaFold predictable model and are colored on the basis of Qres value. The black rectangle highlights the region of association between FtsQ and FtsB. The green colored bond represents the salt bridges and red colored bonds represents the Hydrogen bonds between two proteins. These interactions are conserved in both the structures. (B) PDBsum output shows the interfacial residues for modeled FtsQB from *E. coli*. The red, blue, and orange colored lines represent salt bridge, hydrogen bonds and non-bonded interactions. (C) Front, back, top and bottom view of the AlphaFold prediction model of *E. coli* FtsK'QLBW divisome subcomplex; individual proteins are represented in different colors: FtsK' (blue), FtsQ (red), FtsL (gray), FtsB (orange), and FtsW (yellow). The black circles at the C-terminus highlights the binding regions between FtsQBL and towards N-terminus highlights those between FtsLW. (D) AlphaFold prediction model of the protein complex of FtsQ (Periplasmic domain), FtsL, FtsB, and FtsW in *V. cholerae*, *C. trachomatis*, *M. leprae*, and *M. ulcerans*. The black circle highlights FtsQBL and blue circle FtsLW binding regions respectively.



superimposed on to the crystal structure and was colored on the basis of Qres value (Qres score = 0.85) as shown in Figure 2A. The interactions for FtsQB were extracted from the model (Figure 2B) and compared with the crystal structure (PDB ID: 6H9N) to validate the use of AlphaFold multimer. In the model, interactions between 194 and 256 residues in the periplasmic domain of FtsQ are observed with 52–87 residues of FtsB. Towards the C-terminus, FtsB forms a  $\beta$ -strand that binds to the last  $\beta$ -strand ( $\beta$ -12) of FtsQ to form a continuous  $\beta$ -sheet by antiparallel stacking. FtsB has a loop between the  $\alpha$ -helix and  $\beta$ -strand in the C-terminus, interacting with the Tyr248 of FtsQ. A central hydrophobic patch in FtsQ is formed by Y248 and A253, where FtsB latches onto the FtsQ structure. There are aromatic interactions between FtsQ (Y248) and FtsB (residue Phe84), while FtsB Tyr85 is in close contact with the hydrophobic core of FtsQ formed by residues (L226, L230, V254, and W256). All these interactions involving residues 64–87 of FtsB could be validated from the crystal structure [6H9N] of the complex that containing these residues, though there was minor variation in the distances between the side-chains. The additional interactions seen in the model involve the loop between two sheets of FtsQ (residues 194–197) and one face of the helical region of FtsB (residues 48, 52, 56 and 59). These may be due to the orientation of the proteins and not particularly responsible for binding, as deuterium uptake differences cannot confirm these interactions (Kong et al., 2022).

Interestingly, in the multimer model, FtsL also shows parallel  $\beta$ -sheet stacking when bound to the FtsQ periplasmic domain in the only region where all three proteins are conjoined. The periplasmic domain has only two to three hydrogen bonds between FtsQ and FtsL. These interactions could result from FtsB binding to FtsQ with antiparallel  $\beta$ -sheet packing. Although this extension of the  $\beta$ -sheet stacking seems an elegant utilisation of the extended C-terminus region of FtsL which is otherwise disordered, deuterium uptake studies do not provide sufficient validation for this aspect of the model (Kong et al., 2022).

Cell division proteins FtsB and FtsL form a subcomplex prior to their binding to the FtsQ and other cytoplasmic components of the divisome. The helical transmembrane and putative periplasmic domain portion of the FtsBL subcomplex form an intricate web of hydrophobic contacts and hydrogen bonding that maintain the subcomplex (Condon et al., 2018). Through antiparallel  $\beta$ -sheet packing, the FtsBL subcomplex interacts with FtsQ to produce a 1:1:1 heterocomplex, which can dimerize to form a 2:2:2 complex (Villanelo et al., 2011) without any change to the FtsQBL interface in our model. Leucine residues are found in the proximal periplasmic region of FtsL and the distal periplasmic region of FtsB. Due to the possibility of substituting alternative hydrophobic residues, such as isoleucine or valine, for the leucines that make up the zipper motif, this coiled-coil motif may grow along the

helix. This complex's distal and proximal parts, which lack leucine residues, are stabilized by glutamines, valines, and alanines. The last C-terminus residues of FtsB are free because the FtsL periplasmic domain is shorter.

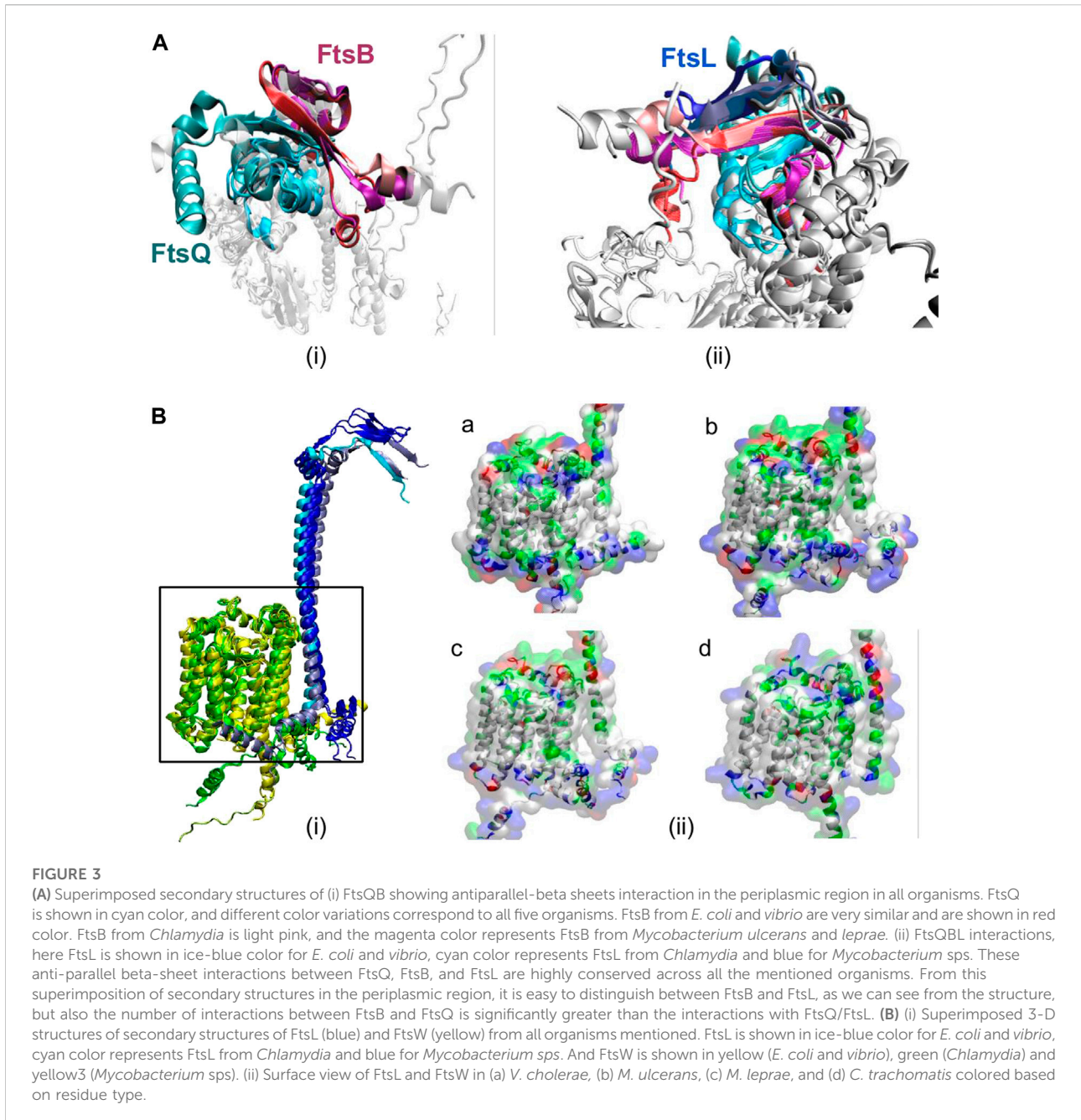
It is believed that FtsW is localized to the septum site by the cytoplasmic region of FtsL; in addition, the predicted multimeric complex also shows interactions with the helical transmembrane and cytoplasmic region of the FtsL (Figure 2C). These proteins interact through two salt bridges in the cytoplasmic domain. Few hydrogen bond interactions occur in the cytoplasm and transmembrane area, while hydrophobic contacts predominate in the transmembrane region. Hence these differential interactions with FtsQ and FtsW can be used to unambiguously assign FtsB and FtsL.

### 3.4 Comparison of FtsQBL interactions in NTD bacteria show similarity to *E. coli*

Multimer model prediction was carried out for FtsQ, FtsL, FtsB, and FtsW in *V. cholerae*, *M. ulcerans*, *M. leprae*, and *C. trachomatis* (Figure 2D). In these organisms, only the periplasmic domain of FtsQ was used, as the disordered cytoplasmic regions of the protein interfered with the FtsBLW interactions. Predicted remote homologs for FtsB and FtsL in *Mycobacterium* spp. also have long disordered N and C termini regions which were excluded from multimer modeling. Based on the monomer structure superimposition for FtsB and FtsL, residue numbers 78–215 of candidate proteins A0PW54 and Q9CD41; and residue numbers 45–180 of A0PTJ5 and Q7AQC6 were considered for multimer modeling. The quality measure for the accuracy of predicted multimer models is the DockQ score. It measures the quality of the interface and gives a score between 0 and 1. A score  $\leq 0.23$  is unacceptable for the model and a score  $\geq 0.8$  is considered a highly accurate model (Basu and Wallner, 2016). The DockQ score for multimer complexes of *V. cholerae*, *M. ulcerans*, *M. leprae* and *C. trachomatis* are 0.69, 0.40, 0.42, and 0.65 respectively. A comparison was done between *E. coli* multimer model and predicted multimer models from all four organisms to delineate interactions among the proteins as mentioned above.

#### 3.4.1 Comparison of intermolecular interactions of FtsQ, FtsB, and FtsL

As seen in the *E. coli* multimer structure, the  $\beta$ -sheet at its C-terminus domain is the point of interaction between the FtsQ molecule and the FtsB/FtsL heterodimer. The C-terminus residues of FtsB (76–88) and the final beta-strand of FtsQ (251–258) are arranged into a twisted  $\beta$ -sheet and are stabilized by multiple hydrogen bonds. The multimer model of *V. cholerae* shares high similarities and is almost identical to *E. coli*, with many interactions between FtsQ and FtsB,



including  $\beta$ -strand and loop formation, conserved in *V. cholerae*.

For *Mycobacterium* sps. and *Chlamydia*, remote homologs of FtsB and FtsL were not distinguished because of their similar structures. These conserved FtsQB secondary structures stacked into antiparallel  $\beta$ -sheet packing in muFtsQ and mlFtsQ were observed with only one of the two candidate proteins that were predicted by the HMM: A0PW54 (muFtsB) and Q9CD41 (mlFtsB) in the periplasmic region of the multimer complex. Secondary structures for the remaining

remote homologs muFtsL (A0PTJ5) and mlFtsL (Q9AQC6) were very similar to FtsL protein. They formed parallel  $\beta$ -sheet interactions with muFtsB and mlFtsB. These proteins muFtsL and mlFtsL also show interactions in the periplasmic domain of FtsQ, but the number of interactions are very few compared to muFtsB and mlFtsB. As seen in multimer structures (Figure 2D), remote homologs muFtsB and mlFtsB in *Mycobacterium* sps. do not interact with FtsW in the cytoplasmic region. These interactions are helpful in distinguishing between the FtsL and FtsB remote homologs

**TABLE 2** This table shows interactions between FtsQ, FtsB, FtsL, and FtsW in all the organisms mentioned in this paper.

Organism	FtsQ:FtsB				FtsQ:FtsL				FtsB:FtsL				FtsL:FtsW			
	IR	SB	HB	NB	IR	SB	HB	NB	IR	SB	HB	NB	IR	SB	HB	NB
<i>E. coli</i>	23:19	5	14	175	9:8	0	5	43	49:48	5	19	285	25:27	2	10	146
<i>V. cholerae</i> (Full Str)	18:19	7	12	169	5:4	0	0	21	47:46	7	16	271	22:26	1	3	107
<i>V. cholerae</i>	19:19	8	14	159	2:1	-	-	7	44:45	7	19	238	21:23	1	4	92
<i>M. ulcerans</i>	16:16	3	12	81	4:4	-	-	10	42:42	3	16	193	6:7	-	1	23
<i>M. leprae</i>	15:13	3	10	63	3:4	-	-	14	43:42	2	14	220	9:10	-	1	36
<i>C. trachomatis</i>	14:13	3	9	90	1:1	-	-	1	44:43	3	17	223	17:15	1	1	78

IR, Interface Residues; SB, Salt Bridge; HB, Hydrogen bond; NB, Non-Bonded Interactions.

in *Mycobacterium* sps. In the multimer complex of *C. trachomatis*, ctFtsB (O84041) forms the same  $\beta$ -strand structure in the C-terminus with an extended helix very similar to the ecFtsB homolog. It also forms antiparallel  $\beta$ -sheet packing with the ctFtsQ and does not interact with FtsW in the cytoplasmic domain. The other predicted remote homolog, ctFtsL (O84273), forms a parallel  $\beta$ -sheet packing with ctFtsB.

Further, the multimer models were superimposed (Figure 3A) with the *E. coli* multimer model based on FtsQ to identify the position of these interacting residues due to their size differences. These superimposed models show proteins muFtsB, mlFtsB, and ctFtsB superimpose with FtsB of *E. coli*. Similarly, muFtsL, mlFtsL, and ctFtsL were superimposed with FtsL of *E. coli*. These multimer models show conserved secondary structure conformations for FtsQ, FtsB, and FtsL remote homologs in *Vibrio cholerae*, *Mycobacterium* sps., and *Chlamydia*, irrespective of their low sequence similarity.

### 3.4.2 Comparison of intermolecular interactions of FtsL and FtsW

As seen in the *E. coli* multimer complex, the cytoplasmic domain of FtsL is crucial for interaction with FtsW and is not needed for interactions with FtsQ and FtsB. From multimer models, it was observed that the remote homologs muFtsL (A0PTJ5), mlFtsL (Q9AQC6), and ctFtsL (O84273) have an extended cytoplasmic tail that binds to FtsW. And muFtsB, mlFtsB, and ctFtsB are slightly away from FtsW protein and do not interact with FtsW.

Structures of the predicted FtsL homolog (muFtsL, mlFtsL, and ctFtsL) and FtsW subcomplex from the complete multimer model were superimposed (Figure 3B) with the *E. coli* FtsLW complex. The functional region between FtsL and FtsW is highly conserved in *V. cholerae* with respect to *E. coli*, as seen in Figure 3B. Similarly, *M. ulcerans* and *M. leprae* also have structural conservation and were superimposed on ecFtsL, but there is an angular shift in the cytoplasmic domain (Figure 3B) which could be a result of long disordered N-terminus. Predicted

remote homolog ctFtsL does not have a long extended cytoplasmic domain, and it interlocks with FtsW very tightly.

### 3.4.3 The number of interactions between FtsQBLW complex proteins are consistent across all organisms

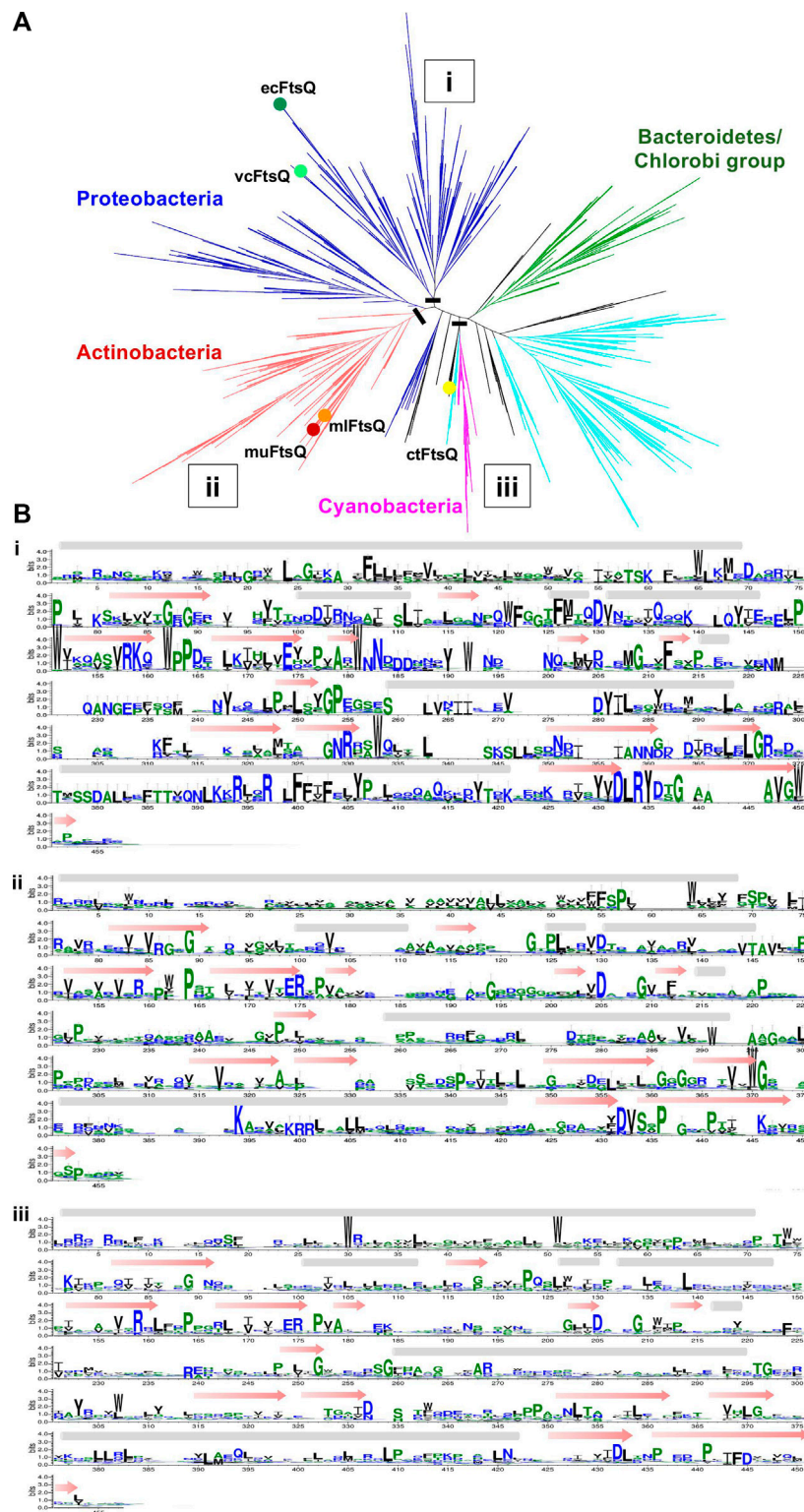
The overall number of interactions between FtsQB, FtsQL, FtsBL, and FtsLW were compared across all the mentioned organisms (Table 2). Multimeric interactions between these proteins were calculated with the help of PDBSum (Laskowski et al., 2018). In *Mycobacterium* sps. and *Chlamydia*, only the periplasmic domain of FtsQ was used for the modeling of FtsQ, FtsB, FtsL, and FtsW as compared to *E. coli* multimer complex where the transmembrane region of FtsQ' and full structure of FtsQ was part of the model. Also, in *Mycobacterium* candidate FtsB (A0PW54, Q9CD41) and FtsL (A0PTJ5, Q7AQC6), only the superimposed functional region with *E. coli* homologs was considered in multimer model building.

Considering major FtsQ and FtsB interactions occur in the periplasmic domain, the number of interactions between FtsQB are almost identical in *E. coli*, and *V. cholera* but are slightly less in *Mycobacterium* sps. and *Chlamydia* (Table 2). The absence of the transmembrane domain of FtsQ in *Mycobacterium* sps. and *Chlamydia* in the model slightly moves the periplasmic domain away from FtsBL while maintaining anti-parallel packing between FtsQB. This was not observed in *V. cholerae* and could result in fewer interactions in *Mycobacterium* sps. and *Chlamydia*.

In *E. coli*, a total of nine and eight interface residues were reported for FtsQ and FtsL. Only 2:2 interfacing residues are present in the periplasmic domain. The number of periplasmic FtsQL interactions in all the mentioned organisms can be considered very similar.

For FtsBL, the number of interacting residues is very close (Table 2). *Mycobacterium* sps. has an extended disordered region on both N and C termini which was excluded from the multimer model prediction.

The interactions in *Mycobacterium* sps. FtsL and FtsW are very few compared to other organisms due to angular shifts in the



**FIGURE 4**

(A) Unrooted cladogram represents the phylogenetic clustering of FtQ from diverse bacterial taxa. The dominant phylum proteobacteria, Firmicutes, actinobacteria, Bacteroidetes/chlorobi group, and cyanobacteria were colored blue, cyan, red, green, and pink respectively. The phyla with less than (0.5%) representation are all colored black. Nodes for the vcFtsQ, muFtsQ, mlFtsQ, and ctFtsQ are highlighted in their respective phyla. Each node shows a divergence of 10% in the phylogenetic tree. (B) Sequence logos representative of the multiple alignments built from different phyla/clades in which the FtQ homologs are present in figure. A black bar shows the location at which tree is cut.

cytoplasmic region of FtsL (Table 2). This angular shift could be because of the long cytoplasmic and periplasmic disordered regions in FtsL, which reduced the interactions between FtsL and FtsW. This is an additional result to distinguish between FtsB and FtsL in *M. ulcerans*, *M. leprae*, and *C. trachomatis*.

### 3.5 Phylogenetic analysis of FtsQ, FtsB, and FtsL homologs

The structures of all orthologs investigated are readily superimposable, with structurally conserved features which can be associated both to their common fold and to their specificity in binding. However, their sequences are not as conserved. Despite this sequence diversity, proteins with a relatively unique fold like FtsQ can be mapped using hidden Markov models. In the case of FtsB and FtsL, their sequence signatures are not sufficiently specific. In order to explore this further, a phylogenetic analysis of the proteins in context with other known orthologs was performed.

All the predicted remote homologs of FtsQ were mapped to COG1589. Multiple sequence alignments built for the FtsQ ortholog cluster have 4226 sequences from 4163 species. The phylogenetic tree in Figure 4 represents the phylogenetic analysis of FtsQ from diverse bacterial taxa. The most abundant phylum in the COG is proteobacteria consisting of almost 38.7% of the full tree, followed by Firmicutes (24.7%), Actinobacteria (17.2%), and Bacteroidetes/Chlorobi (10%). Nodes for the ecFtsQ, vcFtsQ, muFtsQ, mlFtsQ, and ctFtsQ are found to be clustered within their respective phyla. The clustering of ecFtsQ and vcFtsQ in the same clade is indicative of their functional similarity and evolutionary relationship. Similarly, muFtsQ and mlFtsQ are very closely clustered in the clade representing the Actinobacteria phylum. The remote homolog ctFtsQ was detected in the chlamydiae/verrucomicrobia group, which consists of only 0.5% of the phylogenetic tree. The phylum Firmicutes has a separate cluster, but a few sequences from Firmicutes are shown in Figure 4 to be clustering close to the chlamydiae/verrucomicrobia group (0.5%). A sequence logo representative of the multiple alignments (Figure 4) shows that while the proteobacteria clade has sequence patterns that are clearly visible, the clades with mycobacteria FtsQ and clFtsQ have only aromatic and charged residues that stand out. However, the alignment shows few gaps, and the general pattern of hydrophobicity and hydrophilicity is maintained, which can explain the easy identification of orthologs across the bacterial kingdom using hidden Markov models.

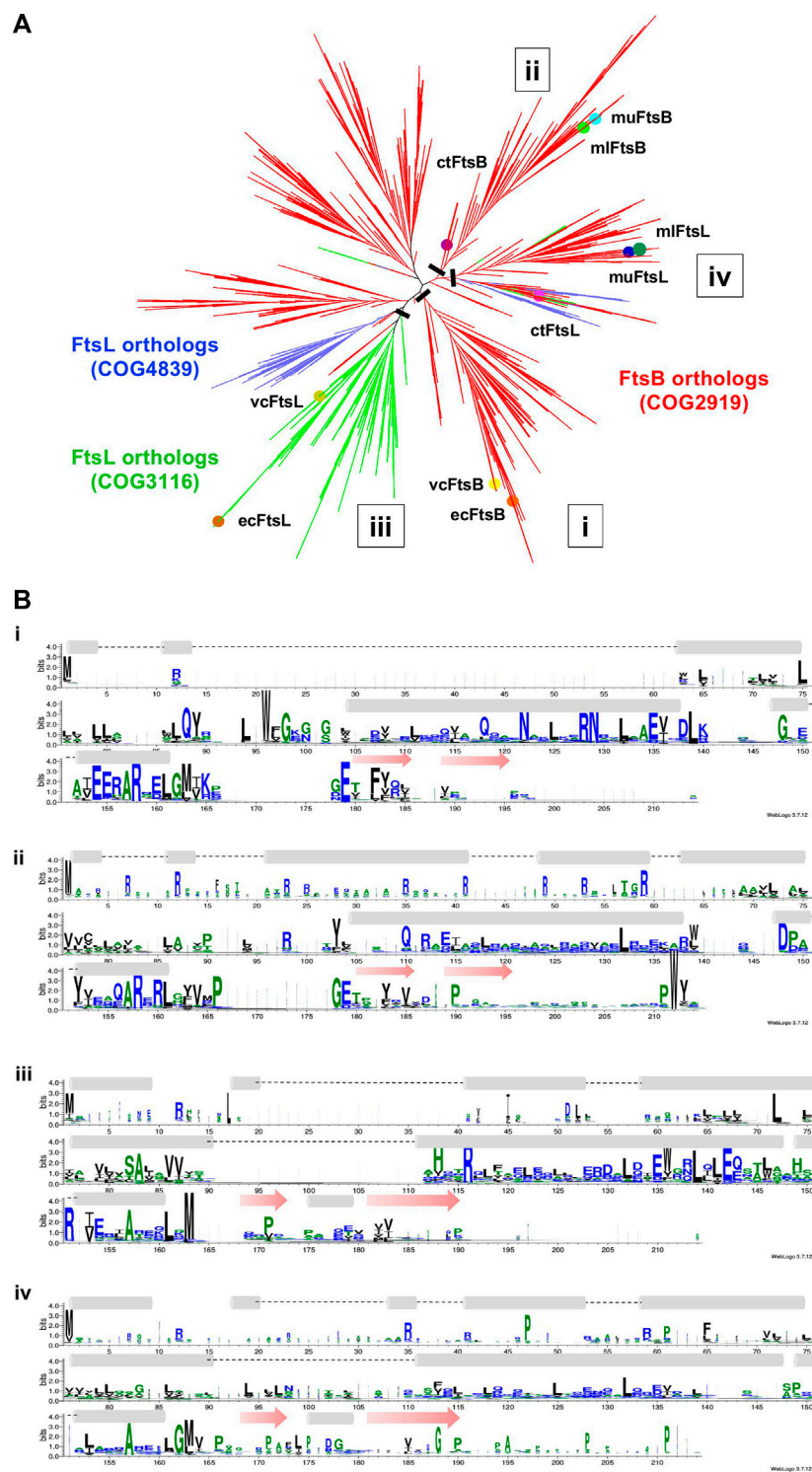
The HMM fails to distinguish FtsB and FtsL unambiguously. In order to investigate why FtsB is occasionally scored higher than the actual ortholog with an HMM prepared from ecFtsL, the clustering patterns for the

sequences of both orthologs were observed together. The FtsB COG2919 has 4644 sequences from 3857 species, out of which 76 are labeled as FtsL. There are two COGs for the FtsL protein: COG3116, which contains 999 sequences from 997 proteobacteria species, and COG4839 which contains 549 sequences from an identical number of species, 98.2% from the Firmicutes phylum. Two proteins ecFtsL and vcFtsL were mapped to COG3116. But muFtsL and mlFtsL did not map to either of them. Surprisingly, muFtsL was found in COG related to penicillin-binding proteins but could be a false positive incorrectly clustered due to the large disordered regions present in these sequences. The other protein mlFtsL maps to FtsB COG, which is understandable as this COG contains many FtsL proteins. As in the case of ctFtsB, ctFtsL is also missing in the database.

All the sequences for FtsB and FtsL from three COGs were taken together, and in addition to this, individual sequences for muFtsL, ctFtsL, and ctFtsB were also added to this dataset. As shown in Figure 5 ecFtsB and vcFtsB are present in the FtsB orthologs cluster, and ecFtsL and vcFtsL are in the same clade in the FtsL orthologs cluster. Homologs for FtsB and FtsL for *Mycobacterium* were found in a different subtree that bifurcates into two clades with muFtsB and mlFtsB present in one clade, muFtsL and mlFtsL present on the other. The ctFtsB is present in the same clade as *Mycobacterium* FtsB but is distant from it. Homolog ctFtsL is present in the clade, which consists of a few taxa from all three COGs. The sequence alignment (Figure 5) for these clades is instructive: only the portion of the common pattern in the amphipathic helix is stacked together. Both proteins have additional domains of unknown function fused to each of the N and C termini of the core regions that interfere with sequence patterns responsible for functional specificity—for FtsL, the N-terminus region that interacts with FtsW and for FtsB, the C-terminus region that interacts with FtsQ.

## 4 Discussion

In this paper, homology-based methods were used in a sequential manner, to identify the homologs of the FtsQBL complex from *E. coli* in four different NTD species that represent both the genome diversity within bacteria phyla and incompletely characterized organisms of potential importance. The pathogen *V. cholerae* has divisive components very similar to *E. coli* and all three homologs were easily identified using sequence-sequence similarity with BLAST. However, as the sequence diversity increases, sequence-sequence comparison methods lose their sensitivity. Increased sensitivity is provided with the profile-sequence comparison HMMER, using profiles created from known orthologs of the query protein, and Profile-Profile (HMM) comparison methods with HH-suite, using profiles built



**FIGURE 5**

(A) Unrooted cladogram represents the phylogenetic analysis of FtsB and FtsL from diverse bacterial taxa. In the phylogenetic tree red subtree represents the FtsB COG2919, green and blue represent the FtsL COG3116 and COG4839. Nodes for FtsB and FtsL homologs are highlighted for all five organisms. (B) Sequence logo for FtsB and FtsL homologs built separately for the two clades in which they are present. A black bar shows the location at which tree is cut.

from a preprocessing step of extracting similar sequences from uniprot20, a version of the Uniprot database with sequence redundancy at 20%. HHsearch method was able to find remote homologs of the FtsQBL complex in *Mycobacterium sps* and *Chlamydia*. The potential remote homologs for FtsQ identified in this study for *M. ulcerans*, *M. leprae*, and *C. trachomatis* are A0PTI5, Q9CCE5, and O84769, respectively. We identified the same protein as ctFtsQ that Ouellette et al., 2015 (Ouellette et al., 2015) reported in their study. Remote homologs of FtsB and FtsL in *Mycobacterium sps*. and *Chlamydia* were not distinguishable with this method, therefore structural modeling of proteins was done using AlphaFold to resolve the ambiguity. These homologs were still ambiguous due to their similar domain and structural fold. To further characterize their function, their multimeric interactions with FtsQ and FtsW were used to distinguish the orthologs. The combined use of fold and protein-protein interactions could be used to map A0PW54 (muFtsB), Q9CD41 (mlFtsB), and O84041 (ctFtsB) as remote homologs of ecFtsB because of their interactions with the periplasmic domain of FtsQ and A0PTJ5 (muFtsL), Q9AQC6 (mlFtsL), and O84273 (ctFtsL) as remote homologs of ecFtsL due to their cytoplasmic domain interaction with FtsW. BLAST alignment for predicted homologs of FtsB and FtsL in *Mycobacterium sps*. with the *M. smegmatis* homologs) (Wu et al., 2018) provide further validation of our approach. The FtsB in *M. smegmatis* is a fusion protein with a domain of unknown function (DUF501) attached to its C-terminus - which is an immediate neighbor of predicted FtsB homologs in both the *Mycobacterium* species. The DUF501 domain maps to residues 227-388 of msFtsB while the muFtsB and mlFtsB map to 1-197 and 1-214 residues of msFtsB. These patterns made it difficult to differentiate between FtsB and FtsL on the basis of phylogeny. Ouellette et al. (2015) (Ouellette et al., 2015) reported ctFtsL homolog with gene name CT\_271 (UniProt ID: O84273), which is the same protein that we characterized as ctFtsL because of its interactions with FtsW in the *C. trachomatis* multimer protein complex. Multimeric interactions played a very important role in successfully characterizing the FtsB and FtsL in *Mycobacterium sps*. and *Chlamydia*. The proteins identified as ctFtsQ and ctFtsL are identical to those previously reported from the experimental findings validates our hypothesis and methodology. In addition, we found a potential remote homolog for ctFtsB.

The hidden Markov model is a mathematical representation of the multiple alignment of sequences in a gene family. Its efficacy is dependent on the quality of clustering sequences into both phylogenetic relationships and gene families which can be used to generate sequence signatures. Common choices for annotation would be the Panther database - which has gene families curated and clustered from 143 genomes into gene families (Thomas et al., 2022) and Inparanoid (Sonnhammer and Östlund, 2015), which identifies and clusters orthologs from pairwise species comparisons but is more focused on Eukaryotic genomes. The EGGNOG database,

created by non-supervised clustering of sequences from all-versus-all pair-wise local alignments, allows for choice in selecting sequences for a gene family at the level of the complete COG, or segmented use at a finer taxonomic level providing some user control on specificity. Profile-profile methods are commonly used to detect diversified proteins with a common fold, and hence ortholog specificity can be even lower with the use of these methods. The HH-suite programs are distributed with the UniRef30 database, created by clustering sequences from the UniRef database with 30% similarity. The use of this default database generated known false positive hits for FtsB and FtsL, especially from the *Mycobacteria* species, which have a number of proteins containing domains of unknown function. Both these proteins functionally interact by forming a coiled-coil, a common motif in many protein-protein interactions. The results described in this paper use an earlier version of a clustered database, uniprot20, that was more specific.

Exploiting the FtsQBL complex proteins interactions with one another and other proteins may aid in discovering drugs that inhibit bacterial growth because of their role in the divisome assembly. In this paper, we modeled the FtsQBLW multimer complex in *M. ulcerans*, *M. leprae*, and *C. trachomatis* and identified their key interactions to shed light on the mechanism of their binding as well as to identify the areas that should be the focus of inhibitors. The interactions between FtsQ and FtsB play an important role in the formation of the FtsQBL complex because FtsB and FtsL bind to each other and then bind to FtsQ with the help of the FtsQB periplasmic domain, which makes the interactions between FtsQB an excellent target for cell division inhibitors. A previous study (Kureisaite-Ciziene et al., 2018) provides experimental findings about the critical role of Tyr248 in the formation of EcFtsQBL complex that shows that the mutations (Y248W and Y248K) have a dominant-negative effect on the FtsQB binding and function. This residue position is on the loop connecting the last two  $\beta$ -sheets towards the C-terminus. The Tyr248 is highly conserved in proteobacteria but has been replaced with Serine in *Mycobacterium* sp. and with Cysteine in *Chlamydia*. These residues can be exploited to act as a specific drug target for inhibition of the FtsQB subcomplex in respective pathogens.

In this paper, our aim is to predict homologs from model organisms like *E. coli* for evolutionary distant species with high confidence. Sequence diversity prevents the identification of the remote homologs in distant species with traditional homology-based methods like BLAST. The more sensitive methods like Profile-Profile comparison along with structural modeling of proteins using AlphaFold - especially in a multimer complex - can be used to assign a specific function to remote homologs that otherwise cannot be easily annotated from traditional sequence analysis methods. The remote homologs we identified are identical to *in vivo* studies which show that the methodology used is capable of detecting homologs in distantly related species, while providing the scope to directly apply structure comparison techniques to study the ortholog. The application of deep learning has recently been made to directly annotate function from a protein's sequence

(Bileschi et al., 2022), and since been applied to the uniprotKB database with higher accuracy, functionally identifying ctFtsQ and ctFtsL, but is still unable to annotate the FtsB and FtsL homologs from *Mycobacterium* and ctFtsB. This technique provides a faster and more accurate alternative to traditional methods in mapping function to orthologs, but is silent on both sequence and structural features responsible for the proteins function which can be derived from conserved sequence signatures and the proteins structure.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Author contributions

HK designed the project, implemented the methods and drafted the manuscript. AL co-designed the project, standardized methods, and provided resources for their implementation.

## Funding

HK is thankful to UGC (University Grants Commission) for providing fellowship. AL is funded from DST, DBT, and ICMR, Govt. of India.

## References

- Aarsman, M. E. G., Piette, A., Fraipont, C., Vinkenvleugel, T. M. F., Nguyen-Disteche, M., and den Blaauwen, T. (2005). Maturation of the *Escherichia coli* divisome occurs in two steps. *Mol. Microbiol.* 55 (6), 1631–1645. doi:10.1111/J.1365-2958.2005.04502.X
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25 (17), 3389–3402. doi:10.1093/NAR/25.17.3389
- Basu, S., and Wallner, B. (2016). DockQ: A quality measure for protein-protein docking models. *PLOS ONE* 11 (8), e0161879. doi:10.1371/JOURNAL.PONE.0161879
- Bateman, A., Martin, M., Orchard, S., Magrane, M., Agivetova, A., Ahmad, S., et al. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49 (D1), D480–D489. doi:10.1093/NAR/GKAA1100
- Bileschi, M., Belanger, D., Bryant, D., Sanderson, T., Carter, B., Sculley, D., et al. (2022). Using deep learning to annotate the protein universe. *Nat. Biotechnol.* 40 (6), 932–937. doi:10.1038/s41587-021-01179-w
- Buddelmeijer, N., and Beckwith, J. (2004). A complex of the *Escherichia coli* cell division proteins FtsL, FtsB and FtsQ forms independently of its localization to the septal region. *Mol. Microbiol.* 52 (5), 1315–1327. doi:10.1111/J.1365-2958.2004.04044.X
- Choi, Y., Kim, J., Yoon, H. J., Jin, K. S., Ryu, S., and Lee, H. H. (2018). Structural insights into the FtsQ/FtsB/FtsL complex, a key component of the divisome. *Sci. Rep.* 8 (1), 18061. doi:10.1038/S41598-018-36001-2
- Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., et al. (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007–1011. doi:10.1038/35059006
- Condon, S. G. F., Mahbuba, D. A., Armstrong, C. R., Diaz-Vazquez, G., Craven, S. J., LaPointe, L. M., et al. (2018). The FtsLB subcomplex of the bacterial divisome is a tetramer with an uninterrupted FtsL helix linking the transmembrane and periplasmic regions. *J. Biol. Chem.* 293 (5), 1623–1641. doi:10.1074/JBC.RA117.000426
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Res.* 14 (6), 1188–1190. doi:10.1101/GR.849004
- Daniel, R. A., Noirot-Gros, M-F., Noirot, P., and Errington, J. (2006). Multiple interactions between the transmembrane division proteins of *Bacillus subtilis* and the role of FtsL instability in divisome assembly. *J. Bacteriol.* 188 (21), 7396–7404. doi:10.1128/JB.01031-06/ASSET/10D1972A-588E-4A11-BE93-460F020582EA/ASSETS/GRAPHIC/ZJB0210661860005.JPEG
- Daumerie, D., Peters, P., and Savioli, L. (2010). *Working to overcome the global impact of neglected tropical diseases: First WHO report on neglected tropical diseases*. Geneva, Switzerland: World Health Organization, 172.
- den Blaauwen, T., Hamoen, L. W., and Levin, P. A. (2017). The divisome at 25: The road ahead. *Curr. Opin. Microbiol.* 36, 85. doi:10.1016/J.MIB.2017.01.007
- Du, S., and Lutkenhaus, J. (2017). Assembly and activation of the *Escherichia coli* divisome. *Mol. Microbiol.* 105 (2), 177–187. doi:10.1111/MMI.13696
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. (2022). Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. doi:10.04.46303410.1101/2021.10.04.463034
- Faruque, S. M., Albert, M. J., and Mekalanos, J. J. (1998). Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* 62 (4), 1301–1314. doi:10.1128/MMBR.62.4.1301-1314.1998

## Acknowledgments

We acknowledge School of Computational & Integrative Sciences, JNU, New Delhi for providing computational facilities.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1010870/full#supplementary-material>



- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 39 (2), W29–W37. doi:10.1093/NAR/GKR367
- Gola, S., Munder, T., Casonato, S., Manganelli, R., and Vicente, M. (2015). The essential role of SepF in mycobacterial division. *Mol. Microbiol.* 97 (3), 560–576. doi:10.1111/mmi.13050
- Gonzalez, M. D., Akbay, E. A., Boyd, D., and Beckwith, J. (2010). Multiple interaction domains in FtsL, a protein component of the widely conserved bacterial FtsLBQ cell division complex. *J. Bacteriol.* 192 (11), 2757–2768. doi:10.1128/JB.01609-09/SUPPL\_FILE/SUPPLEMENTAL\_TABLE\_S1\_S2.ZIP
- HH-suite User Guide *HH-Suite user Guide*. Available at: <https://github.com/soedinglab/hh-suite/wiki> (Accessed: November 26, 2022).
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47 (D1), D309–D314. doi:10.1093/NAR/GKY1085
- Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: Visual molecular dynamics. *J. Mol. Graph.* 14 (1), 33–38. doi:10.1016/0263-7855(96)00018-5
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi:10.1093/MOLBEV/MST010
- Kealey, A., and Smith, R. (2010). N. Tropical diseases: Infection, modeling, and control. *J. Health Care Poor Underserved* 21 (1), 53–69. doi:10.1353/hpu.0.0270
- Kemege, K. E., Hickey, J. M., Barta, M. L., Wickstrum, J., Balwalli, N., Lovell, S., et al. (2015). *Chlamydia trachomatis* protein CT009 is a structural and functional homolog to the key morphogenesis component RodZ and interacts with division septal plane localized MreB. *Mol. Microbiol.* 95 (3), 365–382. doi:10.1111/MML.12855
- Kong, W.-P., Gong, F., So, P. K., Chen, Y. W., Chan, P., Leung, Y. C., et al. (2022). The structural dynamics of full-length divisome transmembrane proteins FtsQ, FtsB, and FtsL in FtsQBL complex formation. *J. Biol. Chem.* 298, 102235. doi:10.1016/j.jbc.2022.102235
- Kureisaite-Ciziene, D., Varadajan, A., McLaughlin, S. H., Glas, M., Silva, A. M., Luirink, R., et al. (2018). Structural organization of the interaction between the bacterial cell division proteins FTSQ and FTSB. *mBio* 9 (5). doi:10.1128/MBIO.01346-18/SUPPL\_FILE/MB0004184054SF6.PDF
- Lapointe, L. M., Taylor, K. C., Subramaniam, S., Khadria, A., Rayment, I., and Benser, A. (2013). Structural organization of FtsB, a transmembrane protein of the bacterial divisome. *Biochemistry* 52 (15), 2574–2585. doi:10.1021/BI400222R/SUPPL\_FILE/BI400222R\_SI\_003.PDB
- Laskowski, R. A., Jablonska, J., Praveda, L., Varekova, R. S., and Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci.* 27 (1), 129–134. doi:10.1002/PRO.3289
- Mercer, K. L. N., and Weiss, D. S. (2002). The *Escherichia coli* cell division protein FtsW is required to recruit its cognate transpeptidase, FtsI (PBP3), to the division site. *J. Bacteriol.* 184 (4), 904–912. doi:10.1128/JB.184.4.904-912.2002/ASSET/27AE29DF-CA06-45B9-A781-F1A3624DC2AB/ASSETS/GRAPHIC/JB0421198005.JPG
- Mirdita, M., von den Driesch, L., Galiez, C., Martin, M. J., Soding, J., and Steinegger, M. (2017). Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* 45, D170–D176. doi:10.1093/NAR/GKW1081
- Nevers, Y., Jones, T. E. M., Jyothi, D., Yates, B., Ferret, M., Portell-Silva, L., et al. (2022). The Quest for Orthologs orthology benchmark service in 2022. *Nucleic Acids Res.* 50 (W1), W623–W632. doi:10.1093/NAR/GKAC330
- Nichio, B. T. L., Marchaukoski, J. N., and Raittz, R. T. (2017). New tools in orthology analysis: A brief review of promising perspectives. *Front. Genet.* 8, 165. doi:10.3389/fgene.2017.00165
- Ouellette, S. P., Karimova, G., Subtil, A., and Ladant, D. (2012). Chlamydia co-opts the rod shape-determining proteins MreB and Pbp2 for cell division. *Mol. Microbiol.* 85 (1), 164–178. doi:10.1111/J.1365-2958.2012.08100.X
- Ouellette, S. P., Rueden, K. J., AbdelRahman, Y. M., Cox, J. V., and Belland, R. J. (2015). Identification and partial characterization of potential FtsL and FtsQ homologs of Chlamydia. *Front. Microbiol.* 6, 1264. doi:10.3389/FMICB.2015.01264
- Ouellette, S. P., Lee, J., and Cox, J. V. (2020). Division without binary fission: Cell division in the FtsZ-less Chlamydia. *J. Bacteriol.* 202 (17), 002522–20. doi:10.1128/JB.00252-20
- Park, K. T., Du, S., and Lutkenhaus, J. (2020). Essential role for ftsI in activation of septal peptidoglycan synthesis. *mBio* 11 (6), 1–17. doi:10.1128/MBIO.03012-20/SUPPL\_FILE/MBIO.03012-20-ST001.DOCX
- Pastoret, S., Fraipont, C., den Blaauwen, T., Wolf, B., Aarsman, M. E. G., Piette, A., et al. (2004). Functional analysis of the cell division protein FtsW of *Escherichia coli*. *J. Bacteriol.* 186 (24), 8370–8379. doi:10.1128/JB.186.24.8370-8379.2004
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLOS ONE* 5 (3), e9490. doi:10.1371/JOURNAL.PONE.0009490
- Remmert, M., Biegert, A., Hauser, A., and Soding, J. (2011). HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* 9, 173–175. doi:10.1038/nmeth.1818
- Roberts, E., Eargle, J., Wright, D., and Luthey-Schulten, Z. (2006). MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinforma.* 7 (1). doi:10.1186/1471-2105-7-382
- Russell, R. B., and Barton, G. J. (1992). Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins* 14 (2), 309–323. doi:10.1002/PROT.340140216
- Söderström, B., and Daley, D. O. (2017). The bacterial divisome: More than a ring? *Curr. Genet.* 63 (2), 161–164. doi:10.1007/s00294-016-0630-2
- Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21 (7), 951–960. doi:10.1093/BIOINFORMATICS/BTI125
- Sonnhammer, E. L. L., and Östlund, G. (2015). InParanoid 8: Orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* 43 (D1), D234–D239. doi:10.1093/NAR/GKU1203
- Steinegger, M., Meier, M., Mirdita, M., Vohringer, H., Haunsberger, S. J., and Soding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinforma.* 20 (1). doi:10.1186/s12859-019-3019-7
- Stephens, R. S., Kalman, S., Lammel, C., Fan, J., MaRathe, R., Aravind, L., et al. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Sci. (New York, N.Y.)* 282 (5389), 754–759. doi:10.1126/SCIENCE.282.5389.754
- Stinear, T. P., Seemann, T., Pidot, S., Frigui, W., Reyssat, G., Garnier, T., et al. (2007). Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res.* 17 (2), 192–200. doi:10.1101/GR.5942807
- Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L. P., and Mi, H. (2022). Panther: Making genome-scale phylogenetics accessible to all. *Protein Sci.* 31 (1), 8–22. doi:10.1002/PRO.4218
- van den Ent, F., Vinkenvleugel, T. M. F., Ind, A., West, P., Veprintsev, D., Nanninga, N., et al. (2008). Structural and mutational analysis of the cell division protein FtsQ. *Mol. Microbiol.* 68 (1), 110–123. doi:10.1111/J.1365-2958.2008.06141.X
- Villanelo, F., Ordenes, A., Brunet, J., Lagos, R., and Monasterio, O. (2011). A model for the *Escherichia coli* FtsB/FtsL/FtsQ cell division complex. *BMC Struct. Biol.* 11 (1), 28–15. doi:10.1186/1472-6807-11-28
- WHO (2022a). Buruli ulcer. Available at: [https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-\(mycobacterium-ulcerans-infection\)](https://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-(mycobacterium-ulcerans-infection)) (Accessed: July 29, 2022).
- WHO (2022b). Leprosy. Available at: <https://www.who.int/news-room/fact-sheets/detail/leprosy> (Accessed: July 29, 2022).
- WHO (2022c). Trachoma. Available at: <https://www.who.int/news-room/fact-sheets/detail/trachoma> (Accessed July 29, 2022).
- Wu, K. J., Zhang, J., Baranowski, C., Leung, V., Rego, E. H., Morita, Y. S., et al. (2018). Characterization of conserved and novel septal factors in *Mycobacterium smegmatis*. *J. Bacteriol.* 200 (6), 006499–17. doi:10.1128/JB.00649-17

## Nomenclature

### Resource identification initiative

**NCBI Datasets** (RRID: SCR\_022569)

**NCBI BLAST** (RRID: SCR\_004870)

**HMMER** (RRID: SCR\_005305)

**Uniprot** (RRID: SCR\_002380)

**eggNOG** (RRID: SCR\_002456)

**VMD** (RRID: SCR\_001820)

**PDBsum** (RRID: SCR\_006511)