Phylogenetic Analysis of FlgJ Protein within Bacteria

Piper Craven

Supervisors: Caroline Puente-Lelièvre, Nick Matzke

School of Biological Sciences

The University of Auckland

Bachelor of Science - Biological Science (Evolution Pathway)

Report approved by Caroline Puente-Lelièvre and Nick Matzke

Student Statement:

The summer research scholarship has furthered my career development by giving me the time, space, financial support, and resources needed to focus exclusively on scientific research. It has also solidified my desire to pursue an advanced degree in biological science. I have gained hands-on experience in an area new to me: protein phylogenetics within bacteria. I have developed the skills necessary to understand evolution and functional biology: analysing protein sequences, constructing phylogenetic trees, and interpreting evolutionary relationships. While working on this project, I learned more about using software to create multiple sequence alignments, find homologs, create phylogenetic trees and run bootstrap calculations. This experience aligns with potential future research and academic careers because it has improved my confidence in tackling complex biological questions and my ability to work with bioinformatics tools.

Abstract:

The bacterial flagella are an incredibly complex system consisting of 20-30 core protein components. FlgJ, a protein important in flagella formation in some bacterial species, has two functions: rod assembly and peptidoglycan hydrolysis. It has been found as a single and a dual-domain protein. The N-terminal domain helps with rod formation, while the C-terminal domain hydrolyses the peptidoglycan layer, allowing the rod to grow. The dual-domain FlgJ has been primarily found in Beta- (β) and Gamma- (γ) proteobacteria, with a fusion event hypothesised to have occurred between 1,4-β-*N*-acetylmuramidase (Acm/muramidase) to the 'core' N-terminal domain of FlgJ. This study used bioinformatics tools, such as iterative homology searches using JackHMMER and phylogenetic inference with IQ-TREE, to investigate how FlgJ has evolved and diversified. We identified three major structurally diverse clades: FlgJ, β-*N*-acetylglucosaminidase, and *N*- acetylmuramoyl-l-alanine amidase. Our results are consistent with the dual-domain hypothesis, as a smaller subset of dual-domain β- and γ-proteobacteria is nested within a larger FlgJ clade primarily composed of single-domain Alpha- (α) proteobacteria. They are also congruent with previous findings showing that FlgJ functions as a β-*N*-acetylglucosaminidase within its catalytic C-terminal rather than a muramidase. Additionally, we report a potential new homolog to FlgJ, *N*-acetylmuramoyl-l-alanine amidase, which functions as a hydrolyse in bacterial cells. This study provides a framework for future research on bacterial motility and flagellar diversity, specifically on rod formation, and early evolution of bacteria.

Introduction:

Bacterial flagella are incredibly complex, multi-component systems that have been the focus of scientific research and debate for decades (Berry & Armitage, 1999; Pallen & Matzke, 2006). The structure of the flagella is a complex of 20-30 protein components comprising three major substructures: (1) the basal body, which traverses the cell wall and receives torque from the motor to transmit to the other structures; (2) the hook, which acts as a universal joint, connecting the basal body to the filaments; and (3) the filament, a helical structure which propels the cell forwards (Fig. 1) (Aizawa & Minamino, 2024; Haiko & Westerlund-Wikström, 2013; Macnab, 2003). Although a basic flagellar architecture is observed across bacteria, there is structural diversity in the location and number of flagella per cell and the number of genes devoted to flagellar formation (Liu & Ochman, 2007; Nakamura & Minamino, 2019). Diversity within the flagella system can reflect environmental adaptations, with motility changing to exploit available resources (Hirano et al., 2001; Nakamura & Minamino, 2019; Wadhwa & Berg, 2022).

FlgJ is a crucial component of the basal body, facilitating both rod assembly and peptidoglycan (PG) hydrolysis (Berg, 2003; Coloma-Rivero et al., 2020; Matsunami et al., 2021; Nambu et al., 1999, 2006; Vollmer et al., 2008). Some FlgJ homologs exist as dual-domain proteins, with an N-terminal domain required for rod assembly and a C-terminal domain needed for peptidoglycan hydrolysis (Herlihey et al., 2014; Hirano et al., 2001; Nambu et al., 1999, 2006; Zhang et al., 2012). Nambu et al. (2006) suggest that a dual-domain FlgJ evolved through a fusion event of a 1,4-β-*N*-acetylmuramidase (Acm/muramidase) with an ancestral single-domain N-terminal (core) FlgJ, forming three homolog groups: canonical dual-domain, non-canonical dual-domain and single-domain type. The canonical dual-domain type, seen primarily in Beta- (β) and Gamma- (γ) proteobacteria, contains the core and the Acm C- terminal domain (Nambu et al., 2006). The non-canonical dual-domain type, found in Delta- (δ) proteobacteria, also retains the core domain but has a M23/M37-family peptidase (Pep) at the C-terminus (Nambu et al., 2006). The single-domain type exists as the core N-terminal of FlgJ (Nambu et al., 2006). Nambu et al. (2006) propose that both dual-domain types have evolved from the single-domain FlgJ, with the Acm or Pep domain fusing at the C-terminus in a common ancestor.

Other homologs exist only as single-domain proteins, having either the C-terminal or N-terminal domain (Nambu et al., 1999, 2006; Zhang et al., 2012). The peptidoglycan hydrolysis properties of FlgJ have often been described as a 1,4-β-*N*-acetylmuramidase (Hirano et al., 2001; Matsunami et al., 2021; Nambu et al., 1999, 2006). However, Herlihey et al. (2014) revealed that during peptidoglycan hydrolysis, FlgJ functions as a β-*N*-acetylglucosaminidase (glucosaminidase) rather than a muramidase (Strating et al., 2012).

Since many motile bacteria are pathogenic and can cause severe infection and disease within various organisms, it is key to understand the molecular basis of flagellar motility and its regulatory mechanisms (Doron & Gorbach, 2008; Ottemann & Miller, 2003; Arenz & Wilson, 2016). This study uses phylogenetic analysis and preliminary protein structure comparisons to investigate the evolutionary relationships between FlgJ and its homologs across bacterial taxa. By re-examining the evolutionary history of FlgJ, our research aims to clarify how FlgJ functions to hydrolyse peptidoglycan and uncover the evolutionary history of the single- and dual-domain FlgJ. These findings can provide new insights into the evolution of bacterial flagella and the molecular mechanisms that shape their structural diversity.

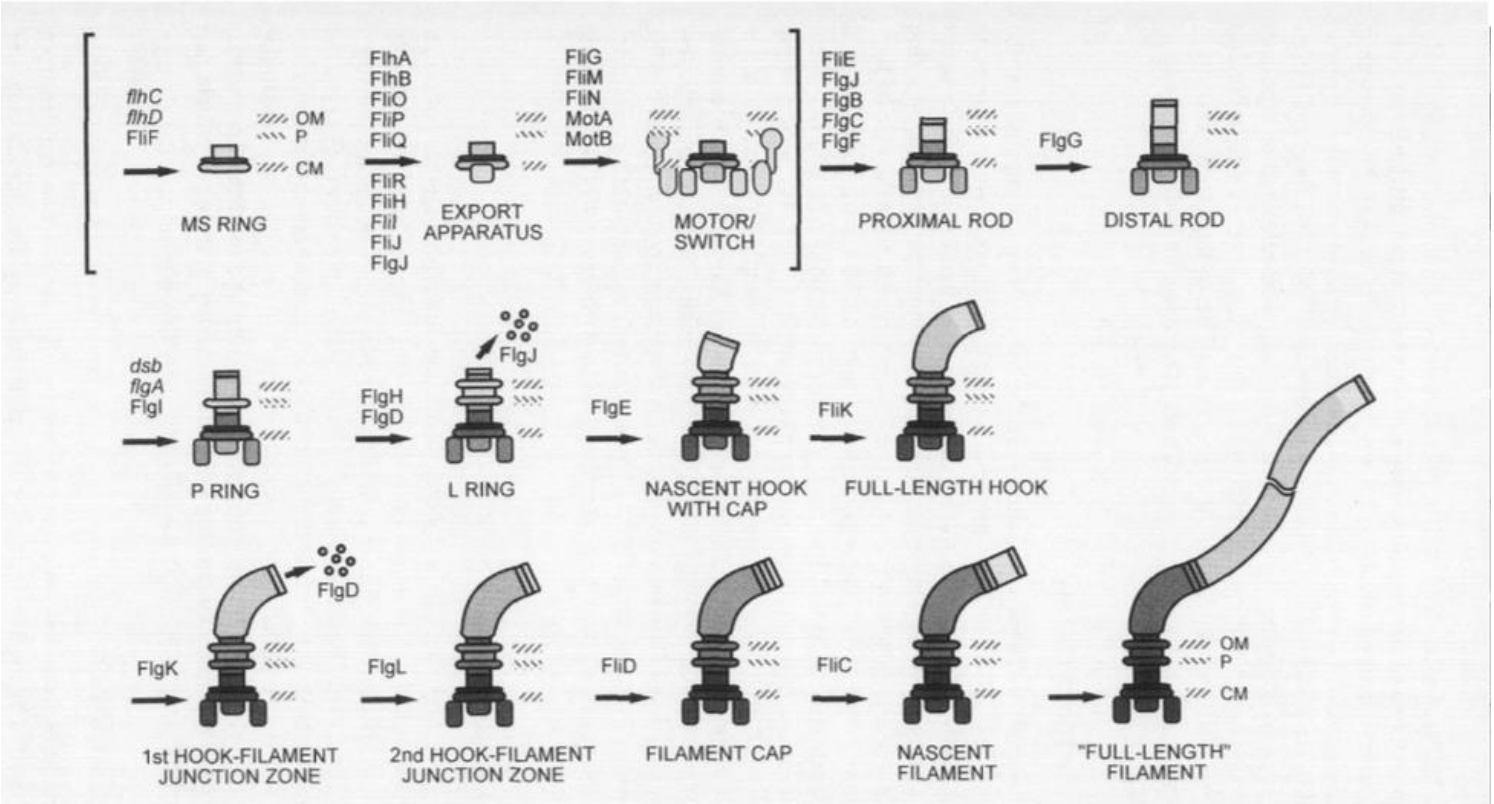

Figure 1. Morphogenetic pathway for the flagellum of *Salmonella*. The brackets indicate substructures that are assembled prior to the utilization of the type III export pathway. The Mot proteins constitute the stator element of the motor and are integral membrane proteins surrounding the MS ring, while FliG and the C ring (FliM and FliN) constitute the rotor/switch of the motor and are peripheral membrane proteins mounted on the MS ring. [FlgJ is found within the export apparatus and during formation of the proximal rod.] Genes (italics) or proteins necessary at each stage are indicated. Figure and figure caption originally published by Macnab (2003).

Materials and Methods:

Remote Homology Search and Sequence Alignment

A set of reference sequences from different bacterial species with known flagellar function (motility) (Appendix 1) was used to perform iterative homology searches using JackHMMER (v. 3.4, 2023; Eddy, 2011) across 193 bacterial genomes (=proteinall in input sequence) (Appendix 2). Reference bacterial species are as follows: *Bacillus subtilis*, *Clostridium difficile*, *Clostridium tetani*, *Enterococcus casseliflavus*, *Escherichia coli K-12*, *Helicobacter pylori*, *Legionella pneumophila*, *Listeria monocytogenes*, *Pseudomonas aeruginosa*, *Salmonella enterica serovar Typhimurium*, and *Vibrio cholerae* (Appendix 1). JackHMMER was run with the default settings of five iterations per reference sequence with the following script:

jackhmmer -o flgJ_hmmer.txt --tblout flgJ_hmmer.csv -A flgJ_hmmer_alignment.fasta FlgJ_ref.txt proteinall.faa

Output sequences were retrieved from the automatically downloaded JackHMMER result files in Stockholm format. These sequences were subsequently converted to FASTA format using a converter from bugaco.com (Bugaco, 2020).

We extracted 1354 homologous sequences from the reference sequences and performed a multiple sequence alignment (MSA) using MUSCLE software integrated within AliView (v. 1.28, 2021, 2014; Larsson, 2014). Duplicate sequences were manually identified and removed based on AliView duplicate sequence identifiers and using ChatGPT (Open AI, 2025) to identify duplicate accession numbers. The final MSA had 352 unique sequences.

Phylogenetic Tree and Structural Protein Construction

Phylogenetic trees were estimated with IQ-TREE (v. 2.3.4, 2024; Trifinopoulos et al., 2016) using the following parameters: automatic substitutional model, ultrafast bootstrap analysis (1000 alignments), 1000 iterations, 0.99 minimum correlation coefficient, Bayes and SH-aLRT branch test (1000 replicates), 0.5 perturbation strength and 100 IQ-TREE stopping rule to search for the best tree (Hoang et al., 2018). The following script was run into IQ-TREE:

Iqtree2 -s i5_deduplicated_copy.fasta -st AA -m TEST -bb 1000 -wbt -alrt 1000 -abayes

The resulting maximum likelihood phylogenetic tree was visualised using FigTree (v. 1.4.4, 2018) and rooted with midpoint rooting.

The PDB file for a β-*N*-acetylglucosaminidase (AlphaFold PSD: AF-M1ZYJ7-F1-v4) and *N*-acetylmuramoyl-l-alanine amidase (Alphafold PSD: AF-D7HEW2-F1-v4) protein structures were found on the AlphaFold protein structure database (Jumper et al., 2021; Varadi et al., 2023; Varadi et al., 2024). The mmCIF file for a FlgJ protein structure (PDB: 5DN4) was found on the RCSB Protein Data Bank (Berman et al., 2000; RCSB.org; Zaloba et al., 2016). Each protein structure was viewed and an image was taken using ChimeraX (v. 1.9, 2024; Meng et al., 2023). The structures were then aligned and viewed together using US-align (Zhang et al., 2022).

Results:

Phylogenetic Analysis

The phylogenetic analysis of 352 FlgJ homologs reveals three major, well-supported clades: FlgJ, β-*N*-acetylglucosaminidase, and *N*-acetylmuramoyl-l-alanine amidase (Fig. 2). The FlgJ clade contains sequences that are known to be associated with flagellar assembly. The β-*N*-acetylglucosaminidase clade appears in the phylogeny as a sister clade to FlgJ (Fig. 2). The *N*-acetylmuramoyl-l-alanine amidase clade exhibits features characteristic of amidase activity and cluster separately from the FlgJ and β-*N*-acetylglucosaminidase clades (Fig. 2). The displayed groupings show that the FlgJ and β-*N*-acetylglucosaminidase clade are more similar due to the FlgJ clade branching from the β-*N*-acetylglucosaminidase clade (Fig. 2). The *N*-acetylmuramoyl-l-alanine amidase clade appears as a diverse but well-defined cluster which displays the most significant divergence from the FlgJ and β-*N*-acetylglucosaminidase clades with the separation occurring at the first divergence from the root (Fig. 2). The sub-clade within the FlgJ clade represents dual-domain FlgJ proteins, species within the β- and γ-proteobacteria class are clustered within this sub-clade (Fig. 2). These sequences form a

distinct group from the single-domain FlgJ proteins, comprised of mainly Alpha- (α) proteobacteria, Epsilon- (ε) proteobacteria, and Spirochaetes (Fig. 2).

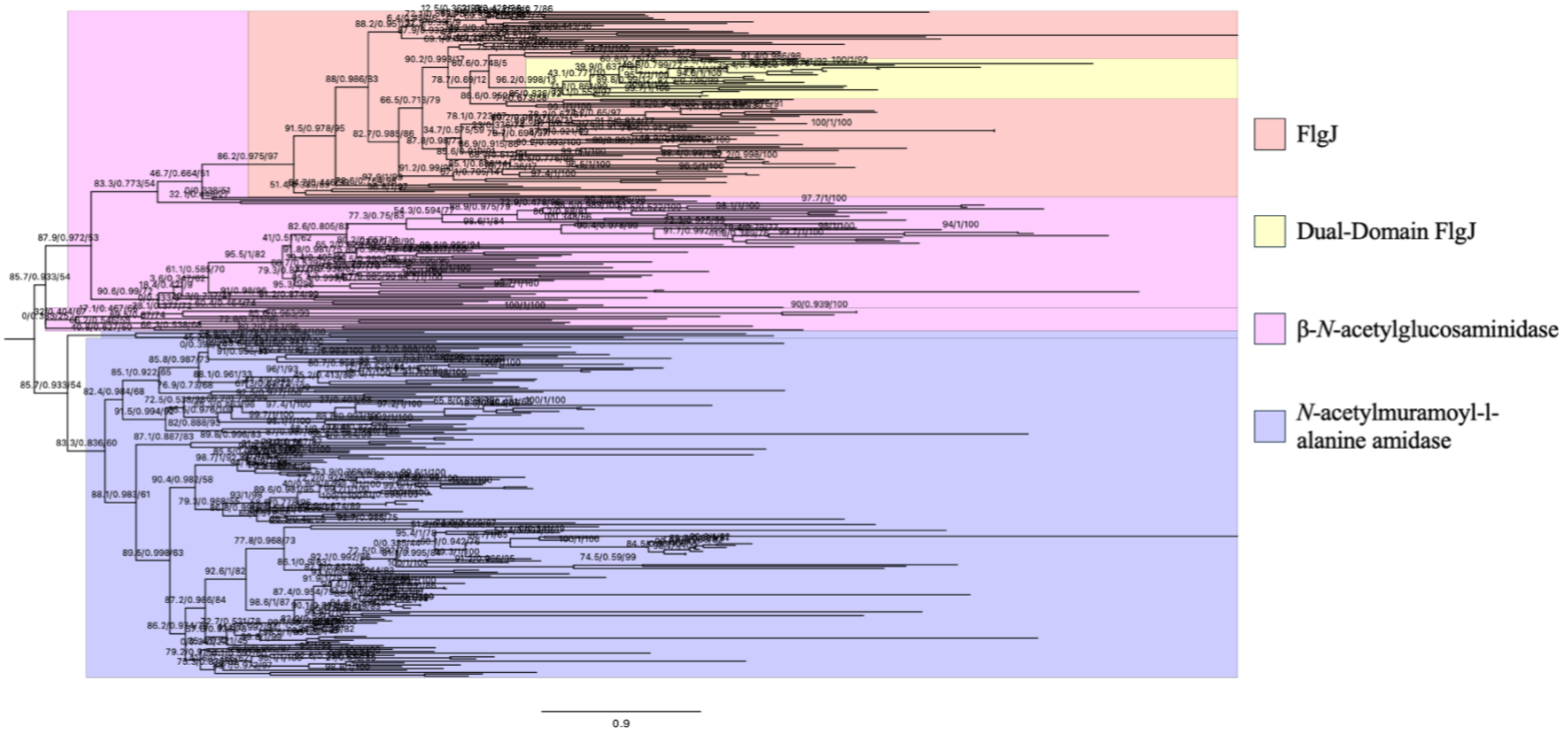

Figure 2. A phylogenetic tree of FlgJ homologs using midpoint rooting was derived from iteratively searching using JackHMMER software (v. 3.4, 2023). Highlighted regions represent clades: FlgJ (orange), Dual-Domain FlgJ (yellow), β-*N*-acetylglucosaminidase (pink), and *N*-acetylmuramoyl-l-alanine amidase (purple). The branch lengths are proportional to the relative amount of time, as indicated by the scale bar (0.9). Branch labels are displayed as SH-aLRT support (%) / aBayes support / ultrafast bootstrap support (%). Data analysis took place at the University of Auckland, New Zealand, between November 2024 and February 2025. Phylogeny was created using IQ-TREE (v. 2.3.4, 2024) and viewed using FigTree (v. 1.4.4, 2018).

Structural 3D Alignment Analysis

The structural alignment between FlgJ, β-*N*-acetylglucosaminidase, and *N*-acetylmuramoyl-l-alanine amidase reveals large degrees of variation in form (Fig. 3). The FlgJ, β-*N*-acetylglucosaminidase, and *N*-acetylmuramoyl-l-alanine amidase multiple structural alignment shows low similarity (TM = 0.235, RMSD = 5.73 Å, SeqID = 0.077) (Fig. 3). Alignment between FlgJ and β-*N*-acetylglucosaminidase showed the most structural difference (TM = 0.257, RMSD = 5.44 Å, SeqID = 0.068) (Fig. 3). β-*N*-acetylglucosaminidase and *N*-acetylmuramoyl-l-alanine amidase have higher structural similarity (TM = 0.282, RMSD = 5.68 Å, SeqID = 0.081), with the alignment between FlgJ and *N*-acetylmuramoyl-l-alanine also having comparable similarity in structures (TM = 0.281, RMSD = 5.22 Å, SeqID = 0.087).

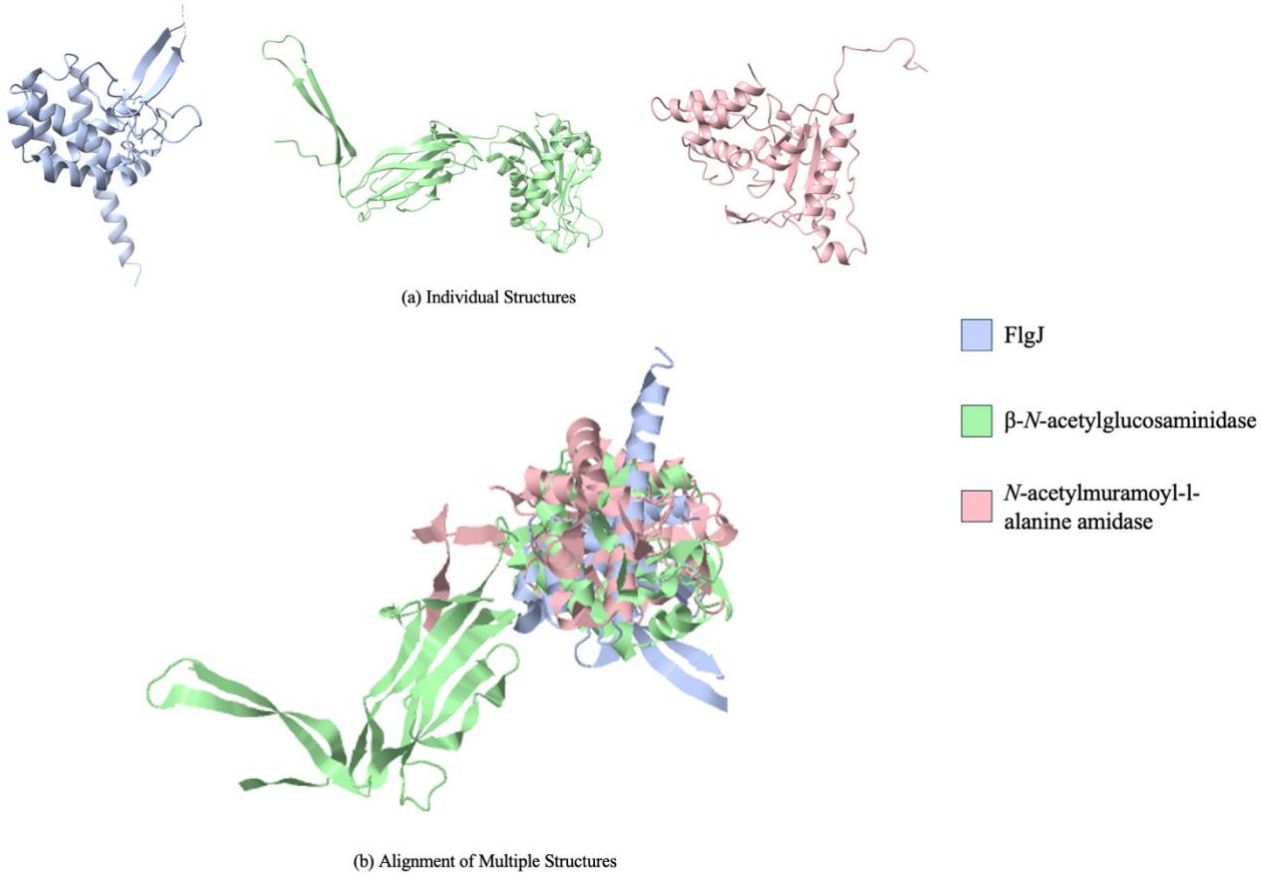

Figure 3. Structural 3D alignment of FlgJ (PDB: 5DN4), β-*N*-acetylglucosaminidase (AlphaFold PSD: AF-M1ZYJ7-F1-v4), and *N*-acetylmuramoyl-l-alanine amidase (Alphafold PSD: AF-D7HEW2-F1-v4). The proteins are displayed in ribbon representation, with each type of protein displayed in different colours (Purple: FlgJ, Green: β-*N*-acetylglucosaminidase, Pink: *N*-acetylmuramoyl-l-alanine amidase). Individual structures (a) were visualised using ChimeraX (v. 1.9, 2024). 3D alignment, analysis and visualisation of multiple structures (b) were performed using US-align (Zhang et al., 2022). All structures were found, aligned, and analysed in February 2025 at the University of Auckland, New Zealand.

Discussion:

Phylogenetic Evidence for Dual-Domain FlgJ

We have identified a FlgJ clade that includes proteins previously annotated by Nambu et al. (2006) as having dual domains. We see three distinct clades within the phylogeny: FlgJ, β-*N*-acetylglucosaminidase and *N*-acetylmuramoyl-l-alanine amidase. FlgJ is comprised of two sub-clades, single-domain and dual-domain. Species within the β- and γ-proteobacteria class, such as *Escherichia coli*, *Salmonella enterica serovar Typhimurium*, and *Vibrio cholerae*, are known to possess dual-domain FlgJ proteins and are seen within the dual-domain sub-clade (Herlihey et al., 2014; Nambu et al., 1999, 2006; Zhang et al., 2012). Bacterial phyla and classes such as α-proteobacteria, ε-proteobacteria and Spirochaetes, thought to possess only the core N-terminal, are all seen clustered in the proposed single-domain clade (Nambu et al.,

1999, 2006; Zhang et al., 2012). This separation may indicate evolutionary pressures shaping the functional role of FlgJ in different bacterial groups (Jayaraman et al., 2022).

β-*N*-acetylglucosaminidase Activity in FlgJ

Our findings are distinguished from the hypothesis proposed by Nambu et al. (2006) that there was a fusion event in a common ancestor with a 'core' N-terminal and 1,4-β-*N*-acetylmuramidase. Our phylogenetic analysis shows that the FlgJ clade is sister to the β-*N*-acetylglucosaminidase clade. This grouping suggests that FlgJ was originally a dual-domain protein, similar to glucosaminidase, which also contains both N-terminal and C-terminal domains, with the C-terminal performing the catalytic functions, much like FlgJ (Herlihey et al., 2014; Nambu et al., 1999, 2006; Rashid et al., 1995). FlgJ has traditionally been classified as a muramidase enzyme, but later studies have shown that FlgJ functions instead as a β-*N*-acetylglucosaminidase (Nambu et al., 1999, 2006; Hirano et al., 2001; Herlihey et al., 2014). This re-classification suggests that the ancestral form of FlgJ may have been a glucosaminidase, and certain bacterial species gradually lost the C-terminal domain due to mutations introducing stop/start codons, rendering the terminal non-functional and leading to a complete loss of the domain (Herlihey et al., 2014; Weiner et al., 2006). Studies have shown that in single-domain proteins, C-terminal loss is more common than N-terminal loss (Weiner et al., 2006). C-terminal domain loss is consistent with our phylogenetic tree, which shows that species within the dual-domain FlgJ clade may have retained both enzymatic functions, while species with only the N-terminal domain likely lost the C-terminal function through independent evolutionary events (Jayaraman et al., 2022). Our confirmation of glucosaminidase as a homolog to FlgJ helps to challenge earlier classifications that primarily associate FlgJ's hydrolase activity with muramidase function (Herlihey et al., 2014; Hirano et al., 2001; Nambu et al., 1999, 2006). Our results suggest that FlgJ originally existed as a dual-domain protein with glucosaminidase activity rather than evolving from a single-domain ancestral form (Herlihey et al., 2014; Nambu et al., 1999, 2006).

N-acetylmuramoyl-l-alanine amidase as a FlgJ Homolog

Phylogenetic and structural analyses show that *N*-acetylmuramoyl-l-alanine amidase (Mur*N*Ac – LAAs) is a homolog of FlgJ. In our study, FlgJ, β-*N*-acetylglucosaminidase, and *N*-acetylmuramoyl-l-alanine amidase form three distinct clades. FlgJ and glucosaminidase are more closely related, sharing a more recent common ancestor. Mur*N*Ac- LAAs diverged earlier and are accordingly more distinct.

Mur*N*Ac – LAAs function as peptidoglycan amidase, which work to dissolve the peptidoglycan layer within bacterial cells to aid in cell division and the separation of daughter cells (Foster, 2004; Vollmer et al., 2008). Many bacteria species, such as *E. coli*, have up to five different types of Mur*N*Ac – LAAs within their proteome (Vollmer et al., 2008).

Both β-*N*-acetylglucosaminidase and *N*-acetylmuramoyl-l-alanine amidase act on peptidoglycan, a heteropolymer of alternating glycan strands and peptide chains composed of repeating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic (MurNAc) linked residues (Herlihey et al., 2014; Vollmer et al., 2008). Mur*N*Ac – LAAs dissolve peptidoglycan by cleaving the amine bond between the lactyl group of MurNAc and the α-amino group of L-alanine, leading to hydrolysis of the peptidoglycan-mesh structure (Herlihey et al., 2014; Foster, 2004; Vollmer et al., 2008). This differs from β-*N*-acetylglucosaminidase, which

typically hydrolyses the 1,4-β linkage between GlcNAc and MurNAc residues in PG (Herlihey et al., 2014).

The clustering of the Mur*N*Ac – LAAs clade, along with its distinct peptidoglycan hydrolysis features, suggests that it has diverged significantly from annotated FlgJ and glucosaminidase lineages (Foster, 2004; Herlihey et al., 2014; Jayaraman et al., 2022). Given that PG is constantly remodelled to support cell growth and division, *N*-acetylmuramoyl-l-alanine amidase likely serves as a functional homolog of FlgJ but with a distinct mechanism for performing its function (Herlihey et al., 2014; Jayaraman et al., 2022). Despite differences in how each enzyme hydrolyses peptidoglycan, both homologs are involved in cell wall modification, suggesting that Mur*N*Ac – LAAs could be a functional relative of FlgJ (Vollmer et al., 2008). Furthermore, Bagon et al. (2021) have linked Mur*N*Ac – LLAs to FlgJ, identifying them as orthologs in predicted secretion pathways in *Lactobacillus johnsonii*. While that study does not explore motility or flagella, this earlier connection points to a potential evolutionary relationship between these enzymes that extends beyond their known roles (Bagon et al., 2021).

Structural Divergence Among FlgJ and its Homologs

The analysis shows considerable structural variation across these homologs. FlgJ and β-*N*-acetylglucosaminidase exhibit the most significant structural differences. *N*-acetylmuramoyl-l-alanine amidase shares slightly higher structural similarity with β-*N*-acetylglucosaminidase than with FlgJ. This structural relationship contrasts with the phylogenetic findings, as FlgJ and β-*N*-acetylglucosaminidase form sister clades, with *N*-acetylmuramoyl-l-alanine being the most divergent. The divergence between the three protein structures could further support the argument for substantial divergence between these clades (Williams & Lovell, 2009). As there is a known relationship between sequence and structural similarity, the structural differences coupled with the difference in function between β-*N*-acetylglucosaminidase and *N*-acetylmuramoyl-l-alanine amidase suggests early sequence divergence (Williams & Lovell, 2009). Insertions, deletions and mutational sensitivity can all contribute to more substantial structural differences between the clades (Marcos & Echave, 2020; Williams & Lovell, 2009).

Interpreting structural data from a small dataset like this can be misleading, as even proteins within the same family can display structural variation (Zea et al., 2018). Additionally, the entire β-*N*-acetylglucosaminidase protein structure did not have very high confidence (90-100 pLDDT), with some areas in the beta-sheets displaying only high confidence (75-89 pLDDT). The phylogenetic analysis is based on available sequence data, which may not capture the full diversity of homologs across bacterial species. Finally, the limited sample size for the structural alignment reduces the accuracy of comparative conclusions. Future studies should incorporate larger structural datasets and higher confidence structures to strengthen these findings.

Conclusion:

These results further suggest the existence of a dual-domain and single-domain FlgJ (Nambu et al., 2006). It is also consistent with Herlihey et al. (2014), which demonstrated that FlgJ functions as a β-*N*-acetylglucosaminidase rather than a 1,4-β-*N*-acetylmuramidase, as previously believed (Hirano et al., 2001; Nambu et al., 1999, 2006). We report a potentially new FlgJ homologue: *N*-acetylmuramoyl-l-alanine amidase, which is known to function as a

peptidoglycan hydrolase within bacteria (Foster, 2004; Vollmer et al., 2008). This enzyme is known to act as a peptidoglycan hydrolase in bacteria though its potential link to FlgJ in bacterial motility and flagellar structure remains unexplored (Bagon et al., 2021; Herlihey et al., 2014; Foster, 2004; Vollmer et al., 2008). Future research should incorporate high-confidence structural data within the phylogenetic context provided here to understand further what affects bacterial motility. Knowing more about how bacteria function can provide us insights into the molecular mechanics of protein synthesis within them (Doron & Gorbach, 2008; Ottemann & Miller, 2003). Such addition to our knowledge of how bacteria use proteins and what those proteins do can provide insight into developing new antibiotics, available treatments, and new targets as bacteria adapt their protein synthesis (Arenz & Wilson, 2016).

References:

Aizawa, S.-I., & Minamino, T. (2024). Flagella. In Y.-W. Tang, M. Y. Hindiyeh, D. Liu, A. Sails, P. Spearman, & J.-R. Zhang (Eds.), *Molecular Medical Microbiology* (Third Edition ed., pp. 97-126). Academic Press. https://doi.org/10.1016/B978-0-12-818619-0.00045-9

Arenz, S., & Wilson, D. N. (2016). Bacterial Protein Synthesis as a Target for Antibiotic Inhibition. *Cold Springs Harbor Perspectives in Medicine, 6*(9). doi: 10.1101/cshperspect.a025361

Bagon, B. B., Valeriano V. D. V., Oh J. K., Pajarillo E. A. B., Lee J. Y., & Kang D.-K. (2021). Exoproteome Perspective on the Bile Stress Response of *Lactobacillus johnsonii*. *proteomes, 9*(10). https://doi.org/ 10.3390/proteomes9010010

Berg, H. C. (2003). The Rotary motor of Bacterial Flagella. *Annual Review of Biochemistry, 72*, 19-54. doi: 10.1146/annurev.biochem.72.121801.161737

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research, 28*, 235-242. https://doi.org/10.1093/nar/28.1.235

Berry, R. M., & Armitage, J. P. (1999). The Bacterial Flagella Motor. In R. K. Poole (Ed.), *Advances in Microbial Physiology* (Vol. 41, pp. 291-337). Academic Press. https://doi.org/10.1016/S0065-2911(08)60169-1

Bugaco. (2020). *Bugaco*. Retrieved 2025, from Bugaco.com: https://sequenceconversion.bugaco.com/converter/biology/sequences/?utm_source=bugaco_home&utm_medium=ref&

Coloma-Rivero, R. F., Gómez, L., Alvarez, F., Saitz, W., del Canto, F., Céspedes, S., . . . Oñate, A. A. (2020). The Role of the Flagellar Protein FlgJ in the Virulence of *Brucella abortus*. *Frontiers in Cellular and Infection Microbiology, 10*. https://doi.org/10.3389/fcimb.2020.00178

Doron, S., & Gorbach, S. (2008). Bacterial Infections: Overview. *International Encyclopedia of Public Health*, 273-282. doi: 10.1016/B978-012373960-5.00596-7

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLOS Computational Biology, 7.* doi: 10.1371/journal.pcbi.1002195

Foster, S. J. (2004). 257 - N-Acetylmuramoyl-l-alanine amidase. In *Handbook of Proteolytic Enzymes* (2nd Edition ed., pp. 866-868). Academic Press. https://doi.org/10.1016/B978-0-12-079611-3.50265-2

Haiko, J., & Westerlund-Wikström, B. (2013). The Role of the Bacterial Flagellum in Adhesion and Virulence. *Biology, 2*(4), 1242-1267. doi: 10.3390/biology2041242

Herlihey, F. A., Moynihan, P. J., & Clarke, A. J. (2014). The Essential Protein for Bacterial Flagella Formation FlgJ Functions as a β-*N*-Acetylglucosaminidase*. *Journal of Biological Chemistry, 289*(45), 31029-31042. doi: 10.1074/jbc.M114.603944

Hirano, T., Minamino, T., & Macnab, R. M. (2001). The role in flagellar rod assembly of the N-terminal domain of *Salmonella* FlgJ, a flagellum-specific muramidase. *Journal of Molecular Biology, 312*(2), 359-369. https://doi.org/10.1006/jmbi.2001.4963

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution, 35*, 518-522. https://doi.org/10.1093/molbev/msx281

Jayaraman, V., Toledo-Patiño , S., Noda-García, L., & Laurino, P. (2022). Mechanisms of protein evolution. *Protein Science, 31*(7). https://doi.org/10.1002/pro.4362

Jumper, J., Evans, R., Pritzel, A. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics*, *30*(22), 3276-3278. https://doi.org/10.1093/bioinformatics/btu531

Liu, R., & Ochman, H. (2007). Origins of Flagellar Gene Operons and Secondary Flagellar Systems. *Journal of Bacteriology, 189*(19), 7098-7104. https://doi.org/10.1128/jb.00643-07

Macnab, R. M. (2003). How Bacteria Assemble Flagella. *Annual Review of Microbiology, 57*, 77-100. doi: 10.1146/annurev.micro.57.030502.090832

Marcos, M. L., & Echave, J. (2020). The variation among sites of protein structure divergence is shaped by mutation and scaled by selection . *Current Research in Structural Biology, 2*, 156-163. doi: 10.1016/j.crstbi.2020.08.002

Matsunami, H., Yoon, Y.-H., Imada, K., Namba, K., & Samatey, F. A. (2021). Structure of the bacterial flagellar hook cap provides insights into a hook assembly mechanism. *Communication Biology, 4*. https://doi.org/10.1038/s42003-021-02796-6

Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for structure building and analysis. *Protein Science, 32*(11). doi: 10.1002/pro.4792

Nakamura, S., & Minamino, T. (2019). Flagella-Driven Motility of Bacteria. *Biomolecules, 9*(7), 279. doi: 10.3390/biom9070279

Nambu, T., Inagaki, Y., & Kutsukake, K. (2006). Plasticity of the domain structure in FlgJ, a bacterial protein involved in flagellar rod formation. *Genes & Genetic Systems, 81*, 381-389. https://doi.org/10.1266/ggs.81.381

Nambu, T., Minamino, T., Macnab, R. M., & Kutsukake, K. (1999). Peptidoglycan-Hydrolyzing Activity of the FlgJ Protein, Essential for Flagellar Rod Formation in *Salmonella typhimurium*. *Journal of Bacteriology, 181*(5), 1555-1561. https://doi.org/10.1128/jb.181.5.1555-1561.1999

Ottemann, K. M., & Miller, J. F. (2003). Roles for motility in bacterial-host interactions. *Molecular Microbiology, 24*(6), 1109-1117. https://doi.org/10.1046/j.1365-2958.1997.4281787.x

Pallen, M. J., & Matzke, N. J. (2006). From The Origin of Species to the origin of bacterial flagella. *Nature Reviews Microbiology, 4*, 784-790. doi: 10.1038/nrmicro1493

Rambaut, A. (2018) Figtree ver 1.4.4. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh

Rashid, M. H., Mori, M., & Sekiguchi, J. (1995). Glucosaminidase of *Bacillus subtilis*: cloning, regulation, primary structure and biochemical characterization. *Microbiology*, *141*(10). https://doi.org/10.1099/13500872-141-10-2391

Strating, H., Vandenende, C., & Clarke, A. J. (2012). Changes in peptidoglycan structure and metabolism during differentiation of proteus mirabilis into swarmer cells. *Canadian Journal of Microbiology, 58*(10). http://dx.doi.org/10.1139/w2012-102

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (n.d.). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research, 44*. https://doi.org/10.1093/nar/gkw256

Varadi, M. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research, 50*, 439-444. https://doi.org/10.1093/nar/gkab1061

Varadi, M. (2023). AlphaFold Protein Structures Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research, 52*. doi: 10.1093/nar/gkad1011

Vollmer, W., Joris, B., Charlier, P., & Foster, S. (n.d.). Bacterial peptidoglycan (murein) hydrolases . *FEMS Microbiology Reviews, 32*(2), 259-286. https://doi.org/10.1111/j.1574-6976.2007.00099.x

Wadhwa, N., & Berg, H. C. (2022). Bacterial motility: machinery and mechanisms. *Nature Reviews Microbiology, 20*, 161-173. https://doi.org/10.1038/s41579-021-00626-4

Weiner, J., Beaussart, F., & Bornberg-Bauer, E. (2006). Domain deletions and substitutions in the modular protein evolution. *FEBS J., 273*(9), 2037-2047. doi: 10.1111/j.1742-4658.2006.05220.x

Williams, S. G., & Lovell, S. C. (2009). The Effect of Sequence Evolution on Protein Structural Divergence. *Molecular Biology and Evolution, 26*(5), 1055-1065. https://doi.org/10.1093/molbev/msp020

Zaloba, P., Bailey-Elkin, B. A., Derksen, M., & Mark, B. L. (2016). Structural and Biochemical Insights into the Peptidoglycan Hydrolase Domain of FlgJ from *Salmonella typhimurium*. *PLos ONE, 11*(2). doi:10.1371/journal.pone.0149204

Zea, D. J., Monzon, A. M., Parisi, G., & Marion-Buslje, C. (2018). How is structural divergence related to evolutionary information? *Molecular Phylogenetics and Evolution, 127*, 859-866. https://doi.org/10.1016/j.ympev.2018.06.033

Zhang, C., Shine, M., Pyle, A. M., & Zhang, Y. (2022). US-align: Universal Structure Alignment of Proteins, Nucleic Acids and Macromolecular Complexes. *Nature Mathods, 19*, 1109-1115. doi: 10.1038/s41592-022-01585-1

Zhang, K., Tong, B. A., Liu, J., & Li, C. (2012). A Single-Domain FlgJ Contributes to Flagellar Hook and Filament Formation in the Lyme Disease Spirochete *Borrelia burgdorferi*. *Journal of Bacteriology, 194*(4), 866-874. https://doi.org/10.1128/jb.06341-11

Appendix:

Reference Species List (Appendix 1)

Bacterial Species List (Appendix 2)