ANALYZING HOST IMMUNE RESPONSE TO CAMPYLOBACTER JEJUNI: A BIOINFORMATICS APPROACH TO BIOMARKER AND DRUG TARGET IDENTIFICATION

by

Divya Ramesh

A Project
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Bioinformatics and Computational Biology

Project Advisor:

_____ Dr. Iosif I. Vaisman, Project Advisor
School of Systems Biology

Date: 12/11/2024

Fall Semester 2024
George Mason University
Manassas, VA

Analyzing Host Immune Response to Campylobacter jejuni: A Bioinformatics Approach to Biomarker and Drug Target Identification

A Research Project submitted in partial fulfillment of the requirements for the degree of Bioinformatics and Computational Biology at George Mason University

by

Divya Ramesh
Master of Science
Manipal Insitute of Higher Education, 2022
Bachelor of Science (Hons)
Sri Ramachandra University, 2016

Director: Dr. Iosif I. Vaisman
School of Systems Biology

Fall Semester 2024
George Mason University
Manassas, VA

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT:

ANALYZING HOST IMMUNE RESPONSE TO CAMPYLOBACTER JEJUNI: A BIOINFORMATICS APPROACH TO BIOMARKER AND DRUG TARGET IDENTIFICATION

Divya Ramesh, M.S

George Mason University, 2024

Project Advisor: Dr. Iosif I. Vaisman

Campylobacter jejuni is a significant cause of bacterial gastroenteritis, impacting millions of people globally and leading to considerable public health concerns. In this study, we sought to identify key proteins involved in the host immune response to C. jejuni infection. Through detailed analysis, we identified a set of critical proteins that play a role in the immune response to this pathogen, including ELANE, MMP9, ACTR2, ACTR3, LCN2, MPO, S100A8, S100A9, CXCL8, and TNF. Further exploration of protein interactions was conducted using the STRING database, which revealed additional proteins interacting with those identified in the initial list, such as ARPC5, ARPC4, ARPC3, ARPC2, WAS, WASL, CTSG, LTF, TLR2, and CD44. This expanded network of 20 proteins provided a comprehensive view of the molecular interactions critical to infection response pathways in C. jejuni infection. To address the challenge of antibiotic resistance in C. jejuni infections, the study also cross-referenced the identified proteins with antibiotics. Azithromycin and Erythromycin emerged as effective treatments for C. jejuni, while Metronidazole showed potential for use in combination therapies, highlighting the ongoing need for novel therapeutic strategies.

To further assess the significance of the identified proteins, Random Forest analysis in R was applied to identify and rank the top 30 proteins from the supplementary data. The mean and standard deviations of these proteins were calculated to evaluate their potential as biomarkers. The Random Forest classification model compared the accuracy of the top 10 proteins from the initial GEO/STRING list with the compiled list of 20 proteins, providing insight into their predictive potential for future diagnostic and therapeutic applications. The findings not only contribute to a better understanding of the molecular mechanisms underlying the host's response to infection but also provide a foundation for further exploration of therapeutic agents.

1. INTRODUCTION:

1.1 CAMPYLOBACTER:

Campylobacter is a genus of Gram-negative, microaerophilic bacteria that has emerged as one of the leading causes of foodborne illness worldwide. These spiral-shaped microorganisms, particularly the species Campylobacter jejuni and Campylobacter coli, are responsible for campylobacteriosis, a gastrointestinal infection affecting millions of people annually. The genus Campylobacter belongs to the family Campylobacteraceae and includes several species, with C. jejuni and C. coli being the most clinically relevant in human infections. (Nam et al., 2023)

1.1.1 Historical context:

The history of Campylobacter dates back to 1886 when Theodor Escherich first observed and described spiral-shaped bacteria in the colons of children who died from what he called "cholera infantum." However, it wasn't until 1963 that the genus Campylobacter was officially proposed by Sebald and Véron, separating it from the genus Vibrio. The development of selective growth media in the 1970s significantly advanced the isolation and identification of Campylobacter, leading to a greater understanding of its role in human disease.

1.1.2 Microbiology and characteristics:

Campylobacter species are slender, spiral, or S-shaped rods, approximately 0.2-0.8 μm wide and 0.5-5 μm long. They are motile, with a characteristic corkscrew-like motion facilitated by a single polar flagellum at one or both ends of the cell. These bacteria are microaerophilic, requiring reduced oxygen levels (typically 3-15% O_2) and increased carbon dioxide (3-5% CO_2) for optimal growth. (Maruyama et al., 2024)

Campylobacter species are fastidious organisms with specific growth requirements. They grow best at temperatures between 37°C and 42°C, with 41.5°C being optimal for C. jejuni. This thermophilic nature contributes to their ability to colonize the intestinal tracts of warm-blooded animals, including

humans. They are sensitive to environmental stresses such as drying, heating, freezing, and exposure to oxygen, yet they can survive in moist environments for extended periods.

1.1.3 Genomics and genetic diversity:

The genomes of Campylobacter species are relatively small, ranging from about 1.6 to 2.0 megabase pairs. C. jejuni, the most well-studied species, has a circular chromosome of approximately 1.6-1.7 Mbp. Despite their small genome size, Campylobacter species exhibit considerable genetic diversity, which contributes to their adaptability and virulence. (Zhang et al., 2021)

Horizontal gene transfer, including natural transformation, plays a significant role in the evolution and diversification of Campylobacter. This genetic plasticity allows for the rapid acquisition of new traits, including antibiotic resistance genes, and contributes to the emergence of new strains with varying pathogenic potential.

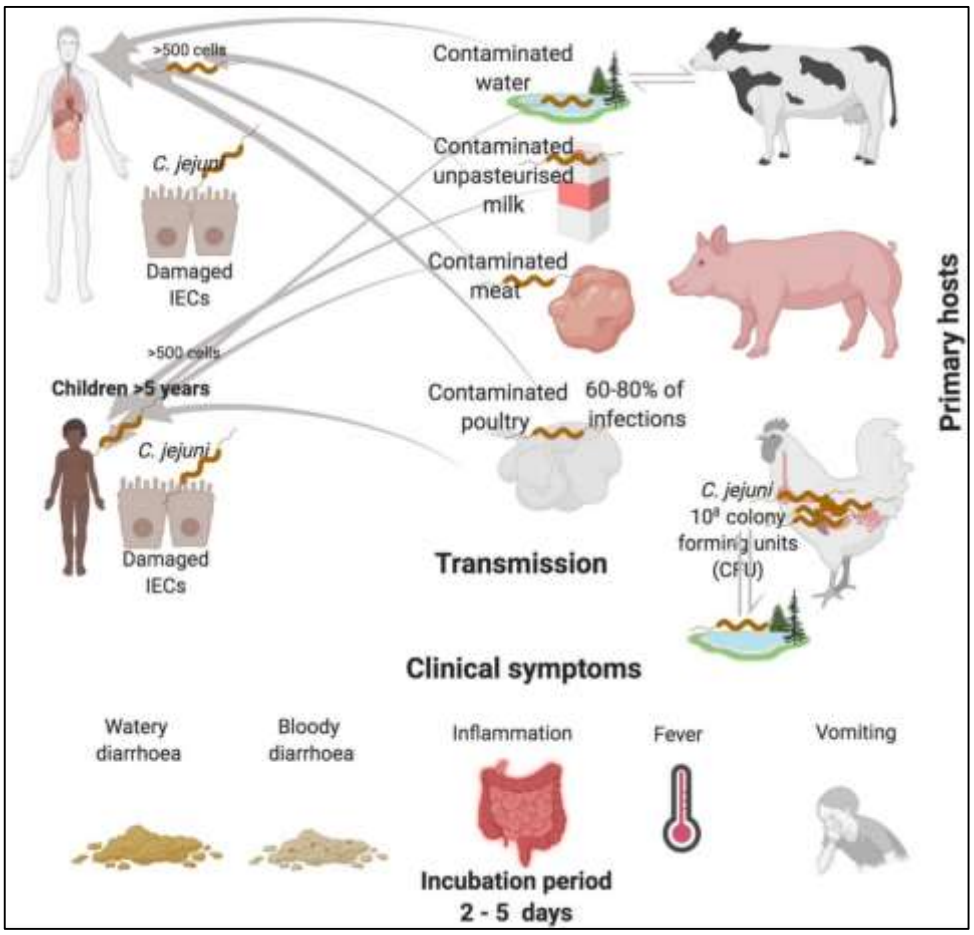


Figure 1: Overview of sources, transmissions and outcomes of C. jejuni infection (Elmi et al., 2021)

1.2 PATHOGENESIS AND VIRULENCE FACTORS:

The pathogenesis of Campylobacter infection involves several steps, including colonization of the intestinal mucosa, adherence to epithelial cells, invasion, and toxin production. Various virulence factors contribute to these processes:

1.2.1 Motility and chemotaxis:

The flagella of Campylobacter are crucial for motility and colonization. The corkscrew-like motion allows the bacteria to penetrate the mucus layer of the intestinal epithelium. Chemotaxis systems guide the bacteria towards favorable environments within the host.

1.2.2 Adhesion and invasion:

Campylobacter possesses several adhesins, including CadF (Campylobacter adhesion to fibronectin), CapA (Campylobacter adhesion protein A), and JlpA (jejuni lipoprotein A), which facilitate attachment to host cells. The ability to invade epithelial cells is an important virulence mechanism, although the exact process is not fully understood.

1.2.3 Toxin production:

While not all strains produce toxins, some Campylobacter isolates can produce cytotoxins, including the cytolethal distending toxin (CDT). CDT causes cell cycle arrest and apoptosis in host cells, contributing to the inflammatory response and tissue damage observed in campylobacteriosis. (Zorba et al., 2020)

1.2.4 Lipooligosaccharide (LOS) and molecular mimicry:

The LOS of Campylobacter plays a role in serum resistance and invasion. Importantly, some LOS structures mimic human gangliosides, which can trigger autoimmune responses leading to conditions such as Guillain-Barré syndrome.

1.3 CLINICAL MANIFESTATIONS:

Campylobacter infections can result in a spectrum of clinical presentations, ranging from asymptomatic carriage to severe, invasive disease.

1.3.1 Acute gastroenteritis:

The most common manifestation of Campylobacter infection is acute gastroenteritis. Symptoms typically include: Diarrhea (often bloody), Abdominal pain and cramping, Fever, Nausea and vomiting (less common). The incubation period is usually 2-5 days, and symptoms typically last for about a week. While most cases are self-limiting, severe cases may require hospitalization, especially in vulnerable populations such as young children, the elderly, and immunocompromised individuals.

1.3.2 Extraintestinal infections:

Although less common, Campylobacter can cause extraintestinal infections, including: Bacteremia, Meningitis, Endocarditis, Septic arthritis and Osteomyelitis. These invasive infections are more likely to occur in immunocompromised patients or those with underlying health conditions.

1.3.3 Post-infectious complications:

Campylobacter infections have been associated with several post-infectious complications:

- Guillain-Barré syndrome (GBS): An autoimmune disorder affecting the peripheral nervous system, occurring in approximately 1 in 1000 Campylobacter infections.
- Reactive arthritis: A form of inflammatory arthritis that can develop following certain infections, including campylobacteriosis.
- Irritable bowel syndrome (IBS): Some studies suggest an increased risk of developing IBS following Campylobacter infection. (Ko et al., 2022)

1.4 EPIDEMIOLOGY AND TRANSMISSION:

Campylobacter is a leading cause of bacterial gastroenteritis worldwide, with an estimated 96 million cases annually. The incidence varies geographically, with higher rates typically observed in developing countries.

1.4.1 Reservoirs and sources of infection:

Campylobacter has a broad host range and can be found in the intestinal tracts of many animals, including: Poultry (chickens, turkeys), Cattle, Sheep, Pigs, Wild birds and Domestic pets (dogs, cats).

1.4.2 Transmission routes:

The primary route of transmission to humans is through the consumption of contaminated food or water. Key sources include undercooked poultry meat, unpasteurized milk and dairy products, contaminated water and cross-contamination during food preparation. Other routes of transmission include direct contact with infected animals and person-to-person transmission (although less common).

1.4.3 Risk factors:

Several factors contribute to the risk of Campylobacter infection:

- Consumption of undercooked poultry
- Travel to regions with poor sanitation
- Occupational exposure (e.g., farm workers, veterinarians)
- Young age (children under 5 years)
- Immunocompromised status

1.5 DIAGNOSIS AND TREATMENT:

1.5.1 Diagnostic methods:

Diagnosis of Campylobacter infection typically involves:

- Stool culture: The gold standard for diagnosis, using selective media and microaerophilic conditions.
- Molecular methods: PCR-based techniques for rapid detection and species identification.
- Antigen detection: Enzyme immunoassays for rapid screening. (Azli et al., 2022)

1.5.2 Treatment approaches:

Most cases of campylobacteriosis are self-limiting and do not require specific treatment beyond supportive care, including fluid and electrolyte replacement and symptomatic relief (e.g., antipyretics, antidiarrheal agents in some cases). Antibiotic therapy is reserved for severe cases, invasive infections, or high-risk patients. Common antibiotics used include macrolides (e.g., azithromycin, erythromycin), fluoroquinolones (e.g., ciprofloxacin) and tetracyclines (e.g., doxycycline). However, the increasing prevalence of antibiotic-resistant Campylobacter strains poses a significant challenge to treatment.

1.6 PUBLIC HEALTH IMPACT AND CONTROL STRATEGIES:

Campylobacter infections represent a substantial public health burden, with significant economic costs related to healthcare expenses, lost productivity, and food industry impacts.

1.6.1 Surveillance and reporting:

Many countries have established surveillance systems to monitor Campylobacter infections. These systems help in detecting outbreaks, identifying trends in incidence and antibiotic resistance and informing public health interventions.

1.6.2 Prevention and control measures:

Efforts to reduce Campylobacter infections focus on multiple levels:

- Farm-level interventions: Improved biosecurity, vaccination of poultry
- Food processing: Enhanced hygiene practices, decontamination methods
- Consumer education: Proper food handling and cooking practices
- Water safety: Ensuring clean water supplies and proper sanitation

1.6.3 Antibiotic resistance:

The emergence of antibiotic-resistant Campylobacter strains is a growing concern. Strategies to address this include surveillance of antibiotic resistance patterns, promoting judicious use of antibiotics in both human medicine and animal husbandry and development of alternative treatment approaches (Kemper & Hensel, 2023)

2. REVIEW OF LITERATURE:

Campylobacter jejuni is one of the leading bacterial pathogens causing gastroenteritis in humans worldwide. Its transmission is primarily through contaminated food and water, leading to symptoms ranging from mild diarrhea to severe illness in vulnerable populations. Understanding the molecular mechanisms of how C. jejuni interacts with the human host is critical to developing effective treatment strategies. Previous research has focused on identifying host immune responses, inflammation pathways, and bacterial virulence factors that contribute to the pathogenesis of C. jejuni infection.

Recent studies utilizing high-throughput genomic databases such as the Gene Expression Omnibus (GEO) have provided new insights into the specific proteins activated during C. jejuni infection. Through data mining in GEO, I filtered relevant datasets and identified key proteins that exhibit differential expression in response to C. jejuni. These proteins serve as important molecular markers for the host's defense mechanism against the pathogen. Building upon these findings, I compiled a list of proteins involved in the infection, which became the basis for further analysis of protein-protein interactions.

Protein-protein interaction networks are essential for understanding the functional relationships between proteins during infection. The STRING database is widely used to explore these interactions by mapping out the neighbouring proteins that interact with known responsive proteins. Leveraging STRING, I identified several additional proteins that consistently appear in the interaction networks of the proteins listed in the GEO dataset. These networked proteins are often central to the infection process and may play crucial roles in mediating the host-pathogen interaction. Identifying common proteins across these networks allowed me to focus on the most relevant molecular players in C. jejuni infection, adding to the existing body of knowledge.

In an effort to explore deeper functional patterns, I utilized the GeneCards database to analyze the identified proteins for shared motifs or patterns. By examining the 20 proteins identified through GEO and STRING, I was able to determine common motifs that may be integral to the response to C. jejuni. This analysis provides a potential framework for understanding how these proteins are regulated during infection.

The current treatment for C. jejuni infections primarily involves the use of antibiotics such as macrolides and fluoroquinolones. However, antibiotic resistance is an increasing concern, and understanding how these drugs interact with bacterial proteins is vital for improving treatment efficacy. In this study, I reviewed the known antibiotics used for C. jejuni treatment, cross-referencing them with the proteins identified in GeneCards. Interestingly, some of the proteins linked to these antibiotics appeared within the 20 proteins from my analysis, suggesting a possible connection between drug targets and host response proteins.

Beyond the known antibiotics, I also identified additional antibiotics that repeatedly occurred within the interaction networks of the identified proteins. These findings suggest alternative therapeutic strategies that could be explored to combat C. jejuni infections, especially in cases where resistance to conventional antibiotics is observed. This insight aligns with recent studies that propose targeting specific host or bacterial proteins as a means to develop more effective treatments against C. jejuni.

Random Forest analysis in R compared the top 10 proteins from the GEO/STRING list with supplementary data, assessing their accuracy as biomarkers. The findings emphasize the importance of targeting host response proteins and bacterial factors to develop more effective treatments for C. jejuni infections.

3. MATERIALS AND METHODS:

***Table 1*:** Tools/Software/Databases used

S.NO	TOOLS/SOFTWARE/DATABASES USED	APPLICATION
1.	NCBI-GEO Dataset	Used to identify key proteins responding to *Campylobacter jejuni* infection by analyzing gene expression datasets
2.	STRING	Utilized to explore protein-protein interaction networks and identify additional proteins interacting with those involved in *C. jejuni* infection
3.	OMIM	Referenced to explore genetic relationships and potential disease associations of proteins involved in *C. jejuni* infection
4.	NIH	Used to gather information on known antibiotics currently employed to treat *C. jejuni* infections and assess their effectiveness
5.	GeneCards	Used to identify common motifs and patterns across proteins and correlate them with antibiotic interactions to explore

		therapeutic targets for *C. jejuni* infection
6.	ML using R – Random Forest	Used to identify the top 10 proteins from the paper and to compare its accuracies with the compiled protein list

3.1 NCBI – GEO Database (Gene Expression Omnibus):

A public repository that archives and freely distributes high-throughput gene expression and other functional genomics data submitted by the scientific community. It's an essential resource for analyzing gene expression patterns across different conditions and species. (Agarwala et al., 2016)

3.2 STRING (Search Tool for the Retrieval of Interacting Genes/Proteins):

A biological database and web resource of known and predicted protein-protein interactions. It integrates both direct (physical) and indirect (functional) associations, providing a comprehensive protein interaction network visualization and analysis platform. (Szklarczyk et al., 2019)

3.3 OMIM (Online Mendelian Inheritance in Man):

A comprehensive, authoritative compendium of human genes and genetic disorders. It focuses on the relationship between phenotype and genotype, containing detailed information about the molecular basis of genetic disorders and inherited traits. (Hamosh et al., 2002)

3.4 NIH (National Institute of Health):

The primary agency of the United States government responsible for biomedical and public health research. Its various databases and resources provide crucial information for biomedical research, clinical trials, and funding opportunities. (National Institutes of Health (NIH), n.d.)

3.5 GeneCards:

An integrative database that provides comprehensive information about all known and predicted human genes. It automatically mines and integrates data from over 150 sources, offering detailed information about gene function, expression, variants, and associated diseases. (Safran et al., 2010)

3.6 ML using R – Random Forest:

Random forest classification in R is a robust machine learning technique that builds multiple decision trees and combines their outputs for more accurate predictions. It is implemented using libraries like randomForest or caret and is well-suited for handling complex datasets with non-linear relationships.

4. WORKFLOW:

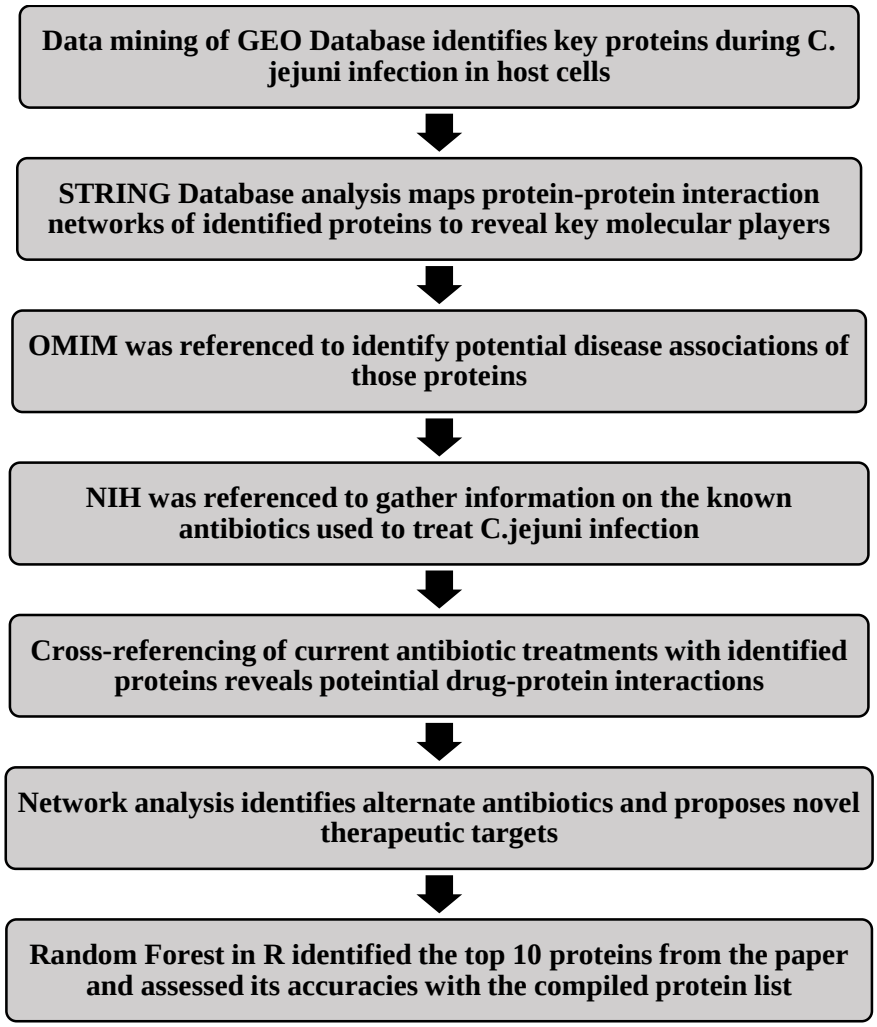


Figure 2: Workflow of the steps followed in identifying novel targets

STEP 1: By mining the GEO database, specific proteins that are activated during the host's immune response to *C. jejuni* were identified, offering insight into the molecular mechanisms of infection. This step helps pinpoint the most relevant proteins showing differential expression in response to the bacterial infection.

STEP 2: Utilizing the STRING database, interaction networks between the identified proteins were mapped, highlighting additional proteins involved in the infection. This analysis helps to explore functional relationships, revealing proteins critical for mediating the host-pathogen interaction.

STEP 3: OMIM database was consulted to cross-reference the proteins with disease associations, enabling the identification of proteins potentially linked to human diseases. This adds further context to the importance of these proteins in broader health-related pathways.

STEP 4: The NIH database provided essential information about current antibiotics used to treat *C. jejuni* infections, such as macrolides and fluoroquinolones. This step ensures that the identified proteins are analyzed for possible interactions with known drug treatments.

STEP 5: By comparing the identified proteins with the current antibiotics, potential drug-protein interactions were discovered. This insight opens up new avenues for understanding how certain proteins may affect or respond to antibiotic treatment.

STEP 6: The protein interaction networks highlighted potential alternative antibiotics and novel therapeutic targets. This finding is crucial in exploring new treatment strategies, especially in cases of antibiotic resistance in *C. jejuni.* (Negretti et al., 2020)

STEP 7: Performed random forest feature selection to identify the top 10 proteins, calculated their mean and standard deviation, and visualized the results through plots. This step is crucial in identifying the essential biomarkers to Campylobacter jejuni.

5. RESULTS:

5.1 DATA MINING STATISTICS:

Table 2 **:** Host response study selected from GEO database analysis

Accession ID	Total number of samples	Title	Proteins identified
GSE147629	78	A porcine ligated loop model reveals new insight into the host immune response against Campylobacter jejuni	Neutrophil elastase, MMP9, ACTR2, ACTR3, LCN2, MPO, S100A8, S100A9, CXCL8, TNF

Following an extensive review of literature focused on Campylobacter jejuni infection and host response mechanisms, the Gene Expression Omnibus (GEO) database was systematically searched to identify relevant experimental datasets. After filtering out various studies explored on Campylobacter jejuni, GSE147629 was selected as the primary dataset due to its robust experimental design using a porcine model and comprehensive analysis of host immune response against C. jejuni infection.

Analysis of this dataset and its supplementary materials revealed ten key proteins that showed significant response during C. jejuni infection: Neutrophil elastase (ELANE), Matrix metallopeptidase 9 (MMP9), Actin-related protein 2 (ACTR2), Actin-related protein 3 (ACTR3), Lipocalin-2 (LCN2), Myeloperoxidase (MPO), S100 calcium-binding protein A8 (S100A8), S100 calcium-binding protein A9 (S100A9), C-X-C motif chemokine ligand 8 (CXCL8), and Tumor necrosis factor (TNF). These proteins represented various aspects of the host immune response, including inflammatory mediators, cytoskeletal regulators, and immune signaling molecules.

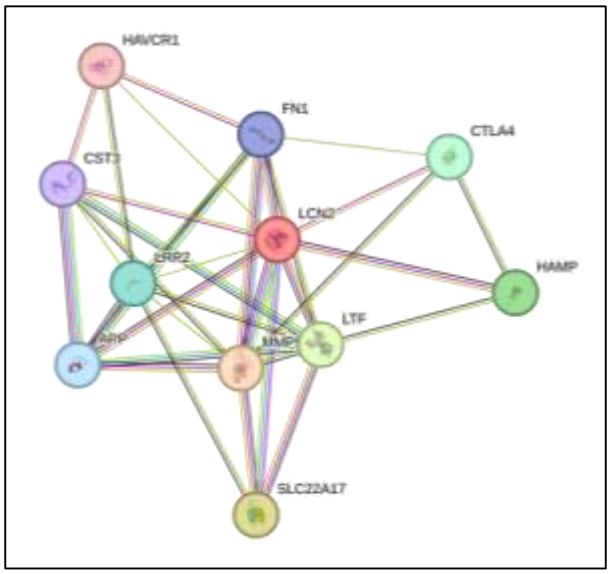
5.2 PROTEIN-PROTEIN INTERACTION:

To better understand the protein interaction networks involved in C. jejuni infection response, a comprehensive analysis was performed using STRING database. Each of the ten proteins identified from the GSE147629 dataset was individually analyzed to identify their neighboring protein interactions.

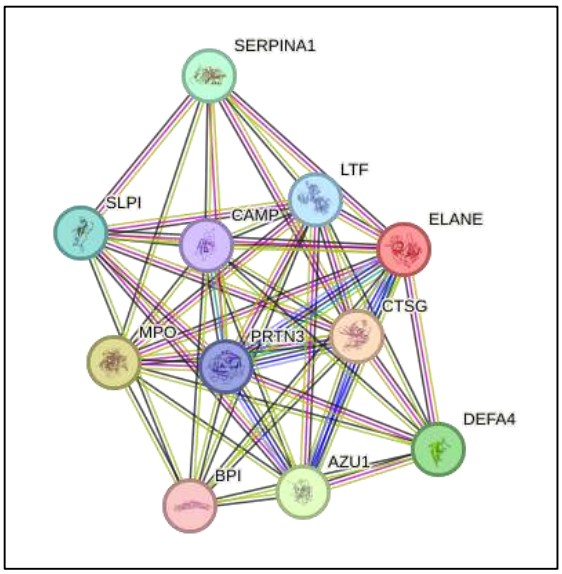
Through this systematic analysis, additional proteins were identified that showed consistent interactions across multiple networks. The proteins **ARPC5, ARPC4, ARPC3, ARPC2, WAS, WASL, CTSG, LTF, TLR2, and CD44** were frequently present in the interaction networks of the initial protein set. These proteins demonstrated high confidence interaction scores (>0.7) and appeared consistently across multiple queries, suggesting their potential importance in the host response to C. jejuni infection.

Given the significance of these consistently interacting proteins, their networks are included in the main text (Figure 2a-2d). This focused set provides insight into the molecular interactions that likely play a critical role in the host's defense against C. jejuni. The remaining protein networks, which show fewer occurrences, are provided in the supplementary material for additional context.

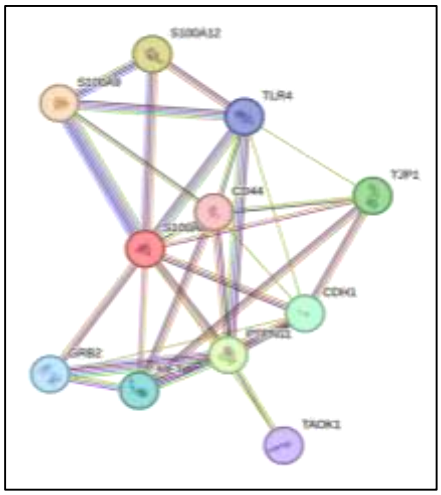
a.)



b.)



c.)



d.)

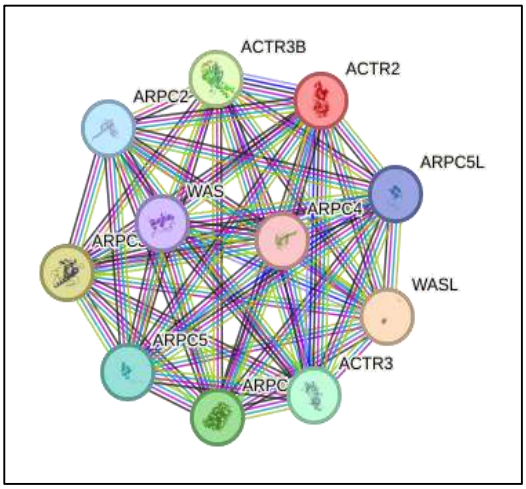


Figure 3 **:** STRING protein-protein interaction showcasing multiple networks related to C. jejuni infection response

The integration of these newly identified proteins with our initial set resulted in a final compilation of 20 proteins. This expanded protein set provided a more comprehensive view of the molecular interaction network involved in C. jejuni infection, including not only the directly responsive proteins but also their key interaction partners that may play crucial roles in the infection response pathway.

5.3 FUNCTIONAL CLASSIFICATION AND DISEASE ASSOCIATIONS OF IDENTIFIED PROTEINS:

Analysis of the 20 identified proteins using GeneCards and OMIM databases revealed distinct functional patterns and disease associations. The proteins clustered into several major functional categories based on their structural motifs and molecular functions.

5.3.1 Cytoskeletal Organization and Regulation

A significant group of proteins (ARPC5, ARPC4, ARPC3, ARPC2, ACTR2, and ACTR3) showed conserved domains associated with actin filament binding and cytoskeletal dynamics. These proteins, particularly components of the ARP-WASP complex, were found to be crucial for phagocytic cup formation and cellular organization. WAS and WASL proteins demonstrated additional functionality in actin nucleation processes, essential for cell motility and immune cell responses.

5.3.2 Immune System Components

Several proteins exhibited characteristic motifs associated with immune function. TLR2, LTF, CTSG, ELANE, LCN2, MPO, S100A8, S100A9, CXCL8, and TNF contained domains specific to innate immune system processes and signaling pathways. Notably, the S100 proteins (S100A8, S100A9) showed distinct calcium-binding domains involved in inflammatory responses.

5.3.3 Enzymatic Activity Patterns

A subset of proteins (CTSG, MMP9, ELANE, and MPO) contained conserved catalytic domains characteristic of endopeptidase activity. These proteins showed motifs essential for protein degradation and tissue remodeling functions.

5.3.4 Disease Associations

OMIM database analysis revealed several key disease associations:

1. Musculoskeletal System:
   - MMP9 and CTSG mutations linked to bone and musculoskeletal disorders
2. Neurological Conditions:
   - ARPC1A, ARPC4, and ARPC5 associated with neurological and genetic disorders
3. Immune System Dysfunction:
   - CTSG, LTF, TLR2, WASL, and WAS mutations connected to various immune-related disorders
   - Particular association with Wiskott-Aldrich Syndrome through WAS, ACTR2, and ACTR3
4. Malignancy:
   - CD44, TLR2, MMP9, WASL, and ARPC2 showed associations with cancer-related conditions
5. Vision:
   - ARPC1A, ARPC5, and ARPC4 linked to eye and vision disorders

The identified proteins showed significant involvement in four key shared pathways: actin dynamics regulation, immune system signaling, Toll-Like Receptor signaling, and matrix metalloproteinase pathways. These findings suggest a complex interplay between cytoskeletal organization, immune response, and tissue remodeling in the context of C. jejuni infection.

5.4 ANTIBIOTIC ANALYSIS AND DRUG RELEVANCE IN C. JEJUNI INFECTION

Following a comprehensive literature review of Campylobacter jejuni treatments in NIH, several known antibiotics were identified that are commonly used to combat the infection. These include **Azithromycin, Erythromycin, Ciprofloxacin, Tetracyclines, Doxycycline, Chloramphenicol, and Fluoroquinolones**. These proteins were cross-referenced against the drugs data in GeneCards by individually analyzing each protein, revealing other potentially relevant antibiotics.

Table 3 **:** Cross-referencing of compiled antibiotics, associated proteins and its function

S.No	Antibiotic	Associated Proteins	Function
1.	Azithromycin	CXCL8	Macrolide antibiotic, first-line treatment for *C. jejuni*, especially in fluoroquinolone-resistant cases
2.	Erythromycin	MMP9, MPO, CXCL8, LTF	Known treatment for *C. jejuni*, effective in multiple networks
3.	Fluoroquinolones	CXCL8, LTF	Broad-spectrum antibiotics, resistance limits first-line use
4.	Tetracycline	None	Broad-spectrum, no association with identified proteins
5.	Doxycycline	None	Broad-spectrum, no association with identified proteins
6.	Chloramphenicol	None	Broad-spectrum, no association with identified proteins
7.	Metronidazole	S100A8, LTF	Effective against anaerobic bacteria, potential for use in combination therapy for *C. jejuni*

8.	Cycloheximide	MMP9, CXCL8, TNF, CD44, TLR2, CTSG	Protein synthesis inhibitor, too toxic for human use
9.	Doxorubicin	MPO, TNF, CD44	Antineoplastic agent, toxicity limits use as an antibiotic
10.	Herbimycin A	MMP9, CXCL8, TNF	Tyrosine kinase inhibitor with antimicrobial properties, not clinically used for bacterial infections

Among these frequently occurring drugs, **Azithromycin**, **Erythromycin** and potentially **Metronidazole** are the most relevant for campylobacter treatment. Azithromycin and Erythromycin is already a known treatment, while Metronidazole has shown some efficacy, especially in combination therapies. The other antibiotics (cycloheximide, doxorubicin, herbimycin A) are either too toxic for routine use or not typically used for bacterial infections in clinical practice. Their presence in this list might suggest some activity against the bacteria or effects on the body's response to infection, but they are not practical treatments for campylobacter (Andersen et al., 2006).

5.5 IDENTIFICATION OF ESSENTIAL BIOMARKERS

A machine learning technique, Random Forest, was employed in R to perform feature selection and identify the top 10 proteins that could serve as potential biomarkers. By applying this technique to the supplementary table data, the top 10 proteins—SERPINB1, S100A12, SERPINB10, OLFM4, GOT2, GPI, TALDO1, LTF, ANXA1, and CYBRD1—were identified. For these proteins, their mean and standard deviation were calculated to further analyze their properties. The results were visualized through various plots, providing insights into their significance and variability. These proteins demonstrate potential as biomarkers, and their identification underscores the utility of machine learning in refining and validating biomarker discovery efforts.

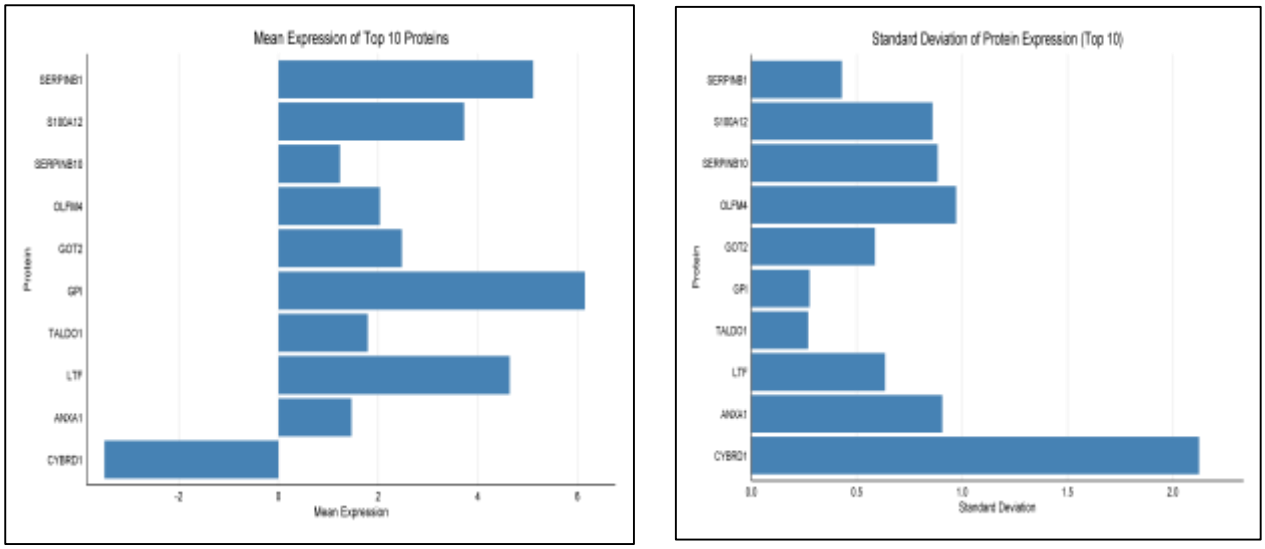


***Figure 4*:** Mean and Standard Deviation of Top 10 proteins based on Random Forest Feature Selection

Random Forest classification was utilized to evaluate and compare the performance of two protein sets: the top 10 proteins derived from the supplementary table and the top 10 proteins selected from a compiled list of 20 proteins. The aim was to assess the classification accuracy of these protein sets and their potential effectiveness as biomarkers.

Table 4: Comparison of AUC for Project and Supplementary Table Proteins

	AUC
Project protein	0.67
Supplementary table proteins	0.75

Results from the analysis revealed that the classification accuracy was relatively low for both sets, indicating challenges in achieving high predictive performance with the selected proteins. However, the supplementary table protein set demonstrated slightly higher accuracy compared to the compiled protein list set. This suggests that the proteins identified from the supplementary table might have a stronger association with the target classification, making them more suitable candidates for further exploration as potential biomarkers. These findings highlight the value of Random Forest classification in comparing the utility of different protein sets in predictive modeling.

6. DISCUSSION:

The findings of this study provide important insights into the molecular mechanisms involved in *Campylobacter jejuni* infections. By identifying proteins through the GEO database and mapping their interactions using STRING, this study highlighted key proteins that participate in host immune responses. These proteins, such as neutrophil elastase and MMP9, are critical to inflammatory pathways that help the body fight against infection (Zeng et al., 2023). The protein-protein interaction analysis not only identified proteins directly involved in the host response but also revealed neighboring proteins that may play critical roles in infection defense mechanisms. This information builds on previous research and adds valuable data to the understanding of *C. jejuni* pathogenesis.

The functional classification of proteins revealed a strong focus on cytoskeletal regulation and immune system components. This classification aligns with *C. jejuni*'s ability to invade host cells and disrupt cellular structures to cause infection. Additionally, the identified disease associations offer further context, as many of these proteins are connected to inflammatory conditions or autoimmune diseases like Guillain-Barré syndrome. These findings underscore the importance of continued exploration into *C. jejuni* infection pathways, especially concerning immune response and cellular disruption.

Moreover, the cross-referencing of antibiotics with identified proteins highlights the ongoing challenge of antibiotic resistance. While erythromycin remains a highly effective treatment, alternative antibiotics such as metronidazole were identified as potential candidates for combination therapies (Gao et al., 2023). This discovery suggests that novel therapeutic strategies can be developed to counteract resistant strains, which is crucial for improving patient outcomes, especially in cases of recurrent or severe infections.

To further evaluate the significance of the identified proteins, Random Forest analysis in R was conducted to identify and rank the top 30 proteins from the supplementary data. The mean and standard deviations of these proteins were calculated to assess their variability and potential as biomarkers. Furthermore, the Random Forest classification model was used to compare the predictive

accuracy of the top 10 proteins derived from the initial GEO/STRING list with those from a compiled list of 20 proteins. This analysis provided valuable insights into the predictive capabilities of these protein sets, offering a foundation for their potential application in future diagnostic and therapeutic strategies.

7. CONCLUSION:

This study offers a comprehensive analysis of the molecular responses to *Campylobacter jejuni* infection and identifies potential therapeutic targets for treatment. The use of databases such as GEO and STRING facilitated the identification of key proteins and their interactions, providing a deeper understanding of the host-pathogen relationship. The insights gained into protein functions, disease associations, and antibiotic interactions contribute to the broader knowledge of *C. jejuni* infections and highlight potential avenues for novel treatment strategies.

Erythromycin and Azithromycin continues to be an effective treatment option, while metronidazole presents an alternative that could be used in combination therapies to combat resistant strains. The findings suggest that targeting both host response proteins and bacterial virulence factors could lead to more effective treatments, reducing the burden of *C. jejuni* infections. This analysis supports the ongoing use of Azithromycin and Erythromycin and also presents Metronidazole as a candidate for further exploration, particularly in combination therapies against C. jejuni infections (Schiaffino et al., 2019). The Machine Learning results show partial overlap with the findings of the paper's statistical analysis, highlighting some alignment with the two approaches

In conclusion, this research underscores the importance of targeting host response proteins, bacterial factors and biomarkers in the development of novel treatments for C. jejuni infections.

8. REFERENCES:

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Charowhas, C., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Funk, K., Geer, L. Y., Gorelenkov, V., … Zbicz, K. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *44*(D1), D7–D19. https://doi.org/10.1093/nar/gkv1290

Andersen, S. R., Shukri, N. M., Boel, J., & Saadbye, P. (2006). Metronidazole resistance in Campylobacter jejuni from poultry meat. *Journal of Food Protection*, *69*(4), 932–934. https://doi.org/10.4315/0362-028X-69.4.932

Azli, B., Razak, M. N., Omar, A. R., Mohd Zain, N. A., Abdul Razak, F., & Nurulfiza, I. (2022). Metagenomics Insights Into the Microbial Diversity and Microbiome Network Analysis on the Heterogeneity of Influent to Effluent Water. *Frontiers in Microbiology*, *13*(April), 1–20. https://doi.org/10.3389/fmicb.2022.779196

Elmi, A., Nasher, F., Dorrell, N., Wren, B., & Gundogdu, O. (2021). Revisiting Campylobacter jejuni Virulence and Fitness Factors: Role in Sensing, Adapting, and Competing. *Frontiers in Cellular and Infection Microbiology*, *10*(February), 1–15. https://doi.org/10.3389/fcimb.2020.607704

Gao, F., Tu, L., Chen, M., Chen, H., Zhang, X., Zhuang, Y., Luo, J., & Chen, M. (2023). Erythromycin resistance of clinical Campylobacter jejuni and Campylobacter coli in Shanghai, China. *Frontiers in Microbiology*, *14*(May). https://doi.org/10.3389/fmicb.2023.1145581

Hamosh, A., Scott, A. F., Amberger, J., Bocchini, C., Valle, D., & McKusick, V. A. (2002). Onlined Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, *30*(1), 52–55. https://doi.org/10.1093/nar/30.1.52

Kemper, L., & Hensel, A. (2023). Campylobacter jejuni: targeting host cells, adhesion, invasion, and survival. *Applied Microbiology and Biotechnology*, *107*(9), 2725–2754. https://doi.org/10.1007/s00253-023-12456-w

Ko, K. K. K., Chng, K. R., & Nagarajan, N. (2022). Metagenomics-enabled microbial surveillance. *Nature Microbiology*, *7*(4), 486–496. https://doi.org/10.1038/s41564-022-01089-w

Maruyama, R., Yasumoto, K., Mizusawa, N., Iijima, M., Yasumoto-Hirose, M., Iguchi, A., Hermawan, O. R., Hosono, T., Takada, R., Song, K. H., Shinjo, R., Watabe, S., & Yasumoto, J. (2024). Metagenomic analysis of the microbial communities and associated network of nitrogen metabolism genes in the Ryukyu limestone aquifer. *Scientific Reports*, *14*(1), 1–14. https://doi.org/10.1038/s41598-024-54614-8

Nam, N. N., Dang, H., Do, K., The, K., Trinh, L., & Lee, N. Y. (2023). Metagenomics : An Effective Approach for Exploring Microbial. *Foods*, *12*, 2140.

Negretti, N. M., Ye, Y., Malavasi, L. M., Pokharel, S. M., Huynh, S., Noh, S., Klima, C. L., Gourley, C. R., Ragle, C. A., Bose, S., Looft, T., Parker, C. T., Clair, G., Adkins, J. N., & Konkel, M. E. (2020). A porcine ligated loop model reveals new insight into the host immune response against Campylobacter jejuni. *Gut Microbes*, *12*(1), 1–25. https://doi.org/10.1080/19490976.2020.1814121

Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., Sirota-Madi, A., Olender, T., Golan, Y., Stelzer, G., Harel, A., & Lancet, D. (2010). GeneCards Version 3: the human gene integrator. *Database : The Journal of Biological Databases and Curation*, *2010*, 1–16. https://doi.org/10.1093/database/baq020

Schiaffino, F., Colston, J. M., Paredes-Olortegui, M., François, R., Pisanic, N., Burga, R., Peñataro-Yori, P., & Kosek, M. N. (2019). Antibiotic resistance of Campylobacter species in a pediatric cohort study. *Antimicrobial Agents and Chemotherapy*, *63*(2), 1–10. https://doi.org/10.1128/AAC.01911-18

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Von Mering, C. (2019). STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Zeng, W., Song, Y., Wang, R., He, R., & Wang, T. (2023). Neutrophil elastase: From mechanisms to therapeutic potential. *Journal of Pharmaceutical Analysis*, *13*(4), 355–366. https://doi.org/10.1016/j.jpha.2022.12.003

Zhang, L., Chen, F. X., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y. J., Hao, H. X., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, *12*(December), 1–15. https://doi.org/10.3389/fmicb.2021.766364

Zorba, G. K., Eleftheriou, T., Engin, İ., Hartsioti, S., & Zenonos, C. (2020). Forensic identification of human remains in Cyprus. *Forensic Science and Humanitarian Action*, 609–623. https://doi.org/10.1002/9781119482062.ch39