A look into the characteristics of *Borrelia garinii*

Bioinformatics
31 October 2025

Carolien Meesters (S5472008),
Fien Habernickel (S5826063),
Lucas Koopmans (S6087159),
Alan Salvador Aguilera Vega (S6509614),
Noa Lamens (S3985830),
Ryan Meuleman (S6462251)

-Abstract-

This case study explores the different properties and aspects of an unknown sequence fragment of DNA. This DNA fragment turned out to be part of the genome of *Borrelia garinii*. Using online tools like NCBI BLAST, ORF finder and KBase certain genes were determined. A literary study provides information about *B.garinii* and its properties and significance, using BLASTp different proteins were analyzed. To research the diagnostics of infection by *B.garinii* an overexpression vector is designed using Benchling.

-Introduction-

This case study analyzed an unknown DNA fragment that was identified as originating from the pathogen *Borrelia garinii* to gain insight into the different properties of the fragment, using online tools such as NCBI BLAST, ORF finder, KBASE and PDB or AlphaFold. Additionally, Benchling was used to design an overexpression vector for diagnostic purposes.

Borrelia garinii is a spiral shaped bacterium (spirochete) with a flagellum between the inner and outer membrane. It contains a small genome and belongs to the *Borrelia* genus. *Borrelia garinii* is **transmitted by ticks** and has an obligate enzootic life cycle [1, 3]. Early stages of infection cause erythema migrans: a red rash which can be accompanied with fever and fatigue. Late stage infection can cause chronic joint inflammation and neuroborreliosis. After a tick bite the *Borrelia* bacteria moves through the skin and into the tissue. Here the bacteria faces the immune system. It turns off the complement system by regulatory protein on its surface, allowing the bacteria to survive. [2] **Chemotaxis** causes *Borrelia* to spread to other parts of the body after the initial bite spot of the tick. The adaptive immune system is helpless because the *Borrelia* bacteria expresses VlsE; a surface protein which repeatedly changes its structure by recombination. Therefore antibodies will not provide protection after or during infection.

-Materials and Methods-

BLAST, ORF Finder and KBase

A BLASTn was performed with the Core nucleotide database in the selected fragment in order to know the bacteria who corresponded to it.
In addition ORF Finder was used to analyze the fragment and find proteins with the following settings: minimal ORF length of 300 nucleotides, genetic code 11, “ATG” and alternative initiation codons including nested ORFs. As well, KBase was applied for gene/protein annotation to obtain CDS and tRNA.

Literary review and phylogeny

With the aim of understanding B. garinii a research was conducted focusing on the bacteria itself, its pathogenicity and virulence factors, Lyme disease, its genetic information and diversity and its significance in (global) health. For the phylogeny SILVA was used to obtain S16 rRNA, Clustal Omega for the alignment and iTOL for the visualization.

Protein and comparative analysis

Three proteins from fragment 6, CheA (fragment 1), CheW (fragment 2) and UvrC (fragment 3), were selected for annotation and analysis. CheA was identified with NCBI smartBLAST and its 3D structure retrieved from PDB; CheW was identified with BLASTp, functionally characterized via NCBI Conserved Domain Database and its structure found in PDB. UvrC was profiled with BLASTp, InterPro and UniProt, also its 3D model obtained from AlphaFold. BLASTp-derived sequences were then aligned in Clustal Omega to get a phylogeny, which was visualized in iTOL and formatted in Sublime Text.

Overexpression Vector Design

Protein annotations were sourced from UniProt, and structures were retrieved from the PDB and AlphaFold. The plasmid carrying the ospC gene was obtained from NCBI, and the expression vector sequence from Thermo Fisher Scientific. Primer design and the cloning strategy were prepared in Benchling to insert ospC into the vector for overexpression of OspC.

-Results-

Scientific name and strain	Total score	Query cover	E value	Accession number
Borrelia garinii SZ	36934	100%	0	CP007564.1
Borrelia garinii NMJW1	36934	100%	0	CP003866.1
Borreliella bavariensis strain JAASAAF1029 chromosome	36928	100%	0	CP075245.1
Borreliella bavariensis strain JANAGAM1369 chromosome	36923	100%	0	CP075235.1
Borreliella bavariensis strain NT24 chromosome	36923	100%	0	CP059009.1

Table 1: BLASTn results of fragment 6

Table 2: Gene and protein annotation results
Showing tRNA, identified protein products not hypothetical (CDS), the products, genomic start positions, strand orientation and sequence lengths obtained from KBase (Prokka) analysis.

Fragment	Product	Start	Strand (+/-)	Length (bp)
1&2	Chemotaxis protein CheW	10,785	-	543
1&2	Protein-glutamate methylesterase/protein-glutamine glutaminase cheB	7,707	-	1,158
3	UDP-N-acetylglucosamine 1-carboxyvinyltransferase murA	1,645	-	1,284
3	Photosystem I assembly protein Ycf3	12634	-	132
1&2	6-phosphogluconate dehydrogenase, NADP(+)-dependent, decarboxylating gndA	14,286	-	1,395
1&2	CTP synthase pyrG	1,788	-	1,602
1&2	Phosphoenolpyruvate-protein phosphotransferase ptsI	1,858	-	1,722
3	UvrABC system protein C uvrC	16077	+	1,812
1&2	Chaperone protein HtpG	16,249	+	1,851
1&2	Chemotaxis protein CheA	9,904	-	2,142

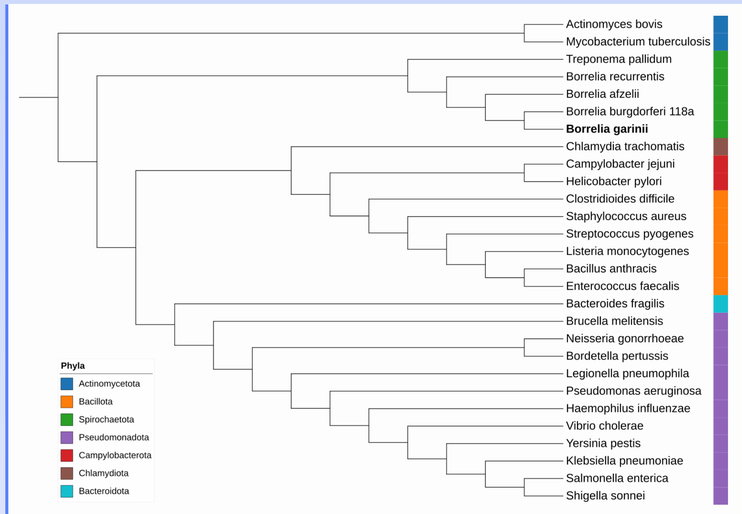


Figure 1: Bacterial phylogeny based on 16S rRNA.
Rectangular phylogenetic tree of representative bacteria inferred from 16S rRNA sequences, leaves are labeled with species names and branches are colored by phylum.

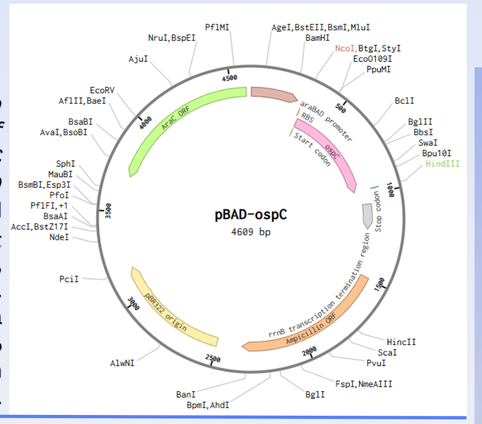


Figure 2- In silico overexpression vector of outer surface protein C (ospC) in pBAD
The RBS is on the vector, and the start codon is located 7–10 nucleotides away from it, on the ospC gene.
The primers used were a forward primer with 31.03% GC and a reverse primer with 38.46% GC content.

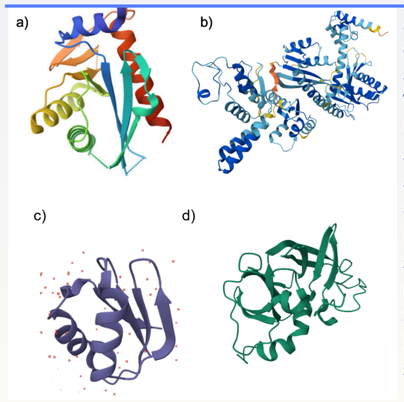


Figure 3 - 3D structures of the proteins
3D structure of Thermatoga maritima UvrC (a) and predicted 3D structure of UvrC (b) 3D structures of CheA and CheW from E. coli, X-Ray diffraction protein structure with a resolution of 2.10 A (c) Solution Nuclear Magnetic Resonance structure (d)

-Conclusion-

The DNA fragment was identified as Borreliella garinii by BLASTn. Its ORFs, genes and proteins were then recognized using NCBI ORFfinder and KBase Prokka.

Literature research examined B. garinii transmission by Eurasian Ixodes ticks, its plasmid rich genome with surface and complement evasion factors enabling infection and immune escape. The review also described Lyme disease presentation, its placement within Lyme borreliae, and how diagnostics (VlsE, OspC) and vaccines (OspA, OspC) are being developed.

Three KBase proteins were analyzed: CheA, a histidine kinase that initiates chemotaxis signaling; CheW, an adaptor that links CheA to chemoreceptors; and UvrC, a nuclease that performs the dual cuts in nucleotide excision repair around damaged DNA.

An in silico pBAD vector was made to overexpress OspC from B. garinii in E. coli TOP10, but because OspC is nontoxic and needed in high amounts, a higher expression system such as pRSET and improved higher GC primers would be more suitable. It requires an experimental test.



-References-

HOW *MYCOPLASMA PNEUMONIAE* MANAGES TO STICK AROUND

university of groningen

Berova I.D. (s5850320), Carl S.A. (s5506506), van der Laan L.D.P. (s5982596), Maire-Richard S. (s5906245), Paingu S. (s5870585), van Veldhuizen R.A. (s6026893)

ABSTRACT

This case study focuses on investigating an unidentified DNA sequence that was blasted and traced back to the pathogen *M. pneumoniae*. To further analyze the genetic and proteomic features of this organisms, KBase, InterPro, AlphaFold, STRING, Pymol and Benchling were utilised.

INTRODUCTION

An unknown DNA fragment was identified as part of the genome from the pathogen *Mycoplasma pneumonia;* the subject of this case study. **Bioinformatics resources** were used to explore the genome and pathogenic gene of the organism and its corresponding protein. In silico analysis was used to design an overexpression vector that could aid in the study of the pathogenic protein and its role in infection.

Mycoplasma pneumoniae is a bacterium that belongs to the genus *Mycoplasma* and the class *Mollicutes*. Its genome has 816,394 bp and 687 genes [1]. *M. pneumoniae* cannot produce a cell wall, the structural support is provided instead by sterols [2]. It is a **parasitic pathogen** that lives on the epithelial surface of its host. It infects only humans, causing atypical community-acquired *pneumonia, pharyngitis* and *tracheobronchitis*. The infection is initiated by cytadherence of the organism to the epithelium by proteins like **P1-adhesin** transmembrane proteins. These bacteria attach to the epithelial membranes of the respiratory tract, damaging the cells, recruiting cytokines and producing hydrogen peroxide and superoxide [3]. Antibodies against *M. pneumoniae* may act against human brain cells and red blood cells, causing anaemia, fever, chest pain and coughing [4].

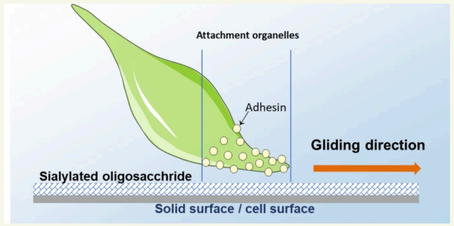


Figure 1: Schematic representation of *M. pneumoniae* and its gliding motility mechanism [5]. The coordinated interaction of adhesin proteins in the bacterium's attachment organelle enables it to bind to sialylated oligosaccharides on a host or solid surface and glide forward in the indicated direction.

MATERIALS & METHODS

An NCBI BLASTx was performed on sequence 3, set 21. KBase was used with Prokka annotation to identify the **genes** present. Literature research was done on *M. pneumoniae* and its infection in humans. For the analysis of the **structure** of adhesin P1, NCBI BLASTP with the database (DB) of the protein data bank (PDB) was performed on the amino acid sequence. AlphaFold was used to predict the protein structure, and PyMOL was used to compare the AlphaFold and PDB structures. To see the difference in the encoded domains of the CDS and PDB protein sequence, InterPro was used. The protein adhesin P1 was searched in the STRING DB. To create an **in silico overexpression** of the mgpA gene, Benchling was used. The primers were designed with PrimerBLAST to isolate the gene from the genome. These primers contained restriction sites needed for the ligation in the pBAD-sfGFP plasmid.

RESULTS

GENE ANALYSIS

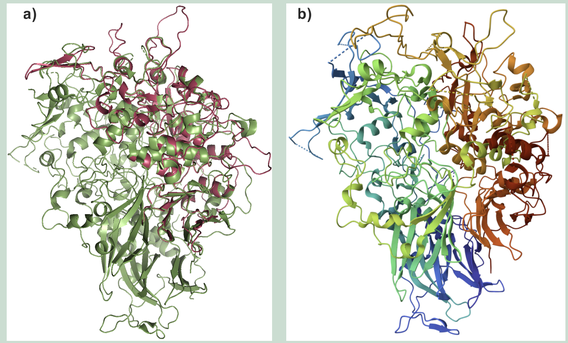
Description	Total Score	Query Cover	E value	Accession
Mycoplasma pneumoniae FH, complete genome	66719	100%	0	CP002077.1
Mycoplasmoides pneumoniae strain 10-1385 chromosome, complete genome	66701	100%	0	CP039784.1
Mycoplasma pneumoniae strain NCTC10119 genome assembly, chromosome: 1	66740	100%	0	LR214945.1
Mycoplasma pneumoniae MAC strain Mac, complete genome	67037	100%	0	CP010550.1
Mycoplasma pneumoniae M129, complete genome	66574	100%	0	U00089.2

Table 1. Raw BLASTN results from the unknown sequence 3 from DNA set 21. All the hits belong to the organism *Mycoplasma pneumoniae*.

Type	Function	Aliases	Start	Strand	Length
CDS	Ribosomal large subunit pseudouridine synthase C	rluC	13,003	minus	930
CDS	Lipoprotein signal peptidase	lspA	13,557	minus	555
CDS	Segregation and condensation protein A	scpA	16,734	plus	1,521
CDS	Segregation and condensation protein B	scpB	18,238	plus	627
CDS	ATP-dependent 6-phosphofructokinase isozyme 1	pfkA	18,884	plus	987
CDS	hypothetical protein		3,441	plus	1,290
CDS	**Adhesin P1**	mgpA	4,833	plus	1,641

Table 2. A summary of the KBase gene annotation of the DNA fragment belonging to *M. pneumoniae*. The gene of interest (adhesin P1, mgaA) is highlighted in bold.

PROTEIN ANALYSIS



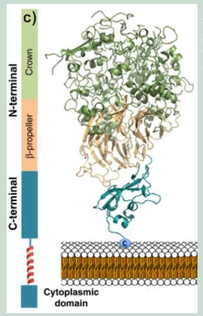


Figure 2: Structural analysis of adhesin P1 in *M. pneumoniae*. NCBI pBLAST (DB: PDB) best hit was Chain A, Adhesin P1 (per. indent. **63.26%**). With InterPro the domains of the PDB structure contained an Adhesin P1 N-terminus domain and C-terminal domain, while the CDS of adhesin P1 only contained a shorter C-terminal domain.

a) PyMOL structural alignment of PDB structure (id: 7BWM) (green) and AlphaFold structure (red).

b) Cryo-EM structure of *M. pneumoniae* Adhesin P1 chain A with resolution of 2.90Å (PDB id: 7BWM). Cartoon representation with color scheme; sequence ID (red= N-terminus, blue= C-terminus).

c) Figure from "*Immunodominant proteins P1 and P40/P90 from human pathogen Mycoplasma pneumoniae*". Crown shown in green comes on top of seven-blade beta-propeller (orange) with the N-terminal domain arranged around it [5].

PROTEIN ANALYSIS

		score
Orf6	Cytadherence protein; MPN142(new), 013(Himmelreich et al, 1996)	0.996
nrnA	DHH family phosphoesterases; Bifunctional enzyme which has both oligoribonuclease and pAp- phosphatase activities.	0.919
hmw1	Cytadherence accessory protein HMW1; component of the cytoskeleton-like scructure which stabalizes the shape of the wall-less Mycoplasma.	0.919
hmw3	Cytadherence accessory protein HMW3; component of the cytoskeleton-like scructure which stabalizes the shape of the wall-less Mycoplasma.	0.918
p30	30K adhesin-related protein; Adhesin necessary for succesful cytadherence and virulence.	0.899
E07_orf175	Hypothetical protein; MPN143(new), 012(Himmelreich et al, 1996)	0.802
P65	Conserved hypothetical protein; ; MPN143(new), 528(Himmelreich et al, 1996)	0.801
hmw2	Cytadherence high molecular weight protein 2; Component of the cytoskeleton-like structure which stabalizes the shape of the wall-less Mcoplasma	0.76
dnaK	DnaK; Acts as a chaperone; Belongs to the heat shock protein 70 family	0.731
E07_orf166	Hypothetical protein; MPN138(new), 017(Himmelreich et al, 1996); see: MPN013; Belongs to the UPF0134 family	0.708

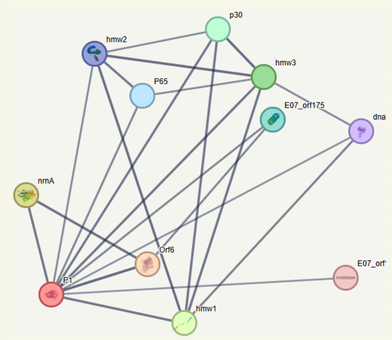


Figure 3: The full STRING network, visualized in STRING. The string is a **prediction** of how well other **proteins partner** with the adhesin P1 protein. The **factors** looked at were: gene neighborhoods, gene fusions, gene co-occurrence, textmining, co-expression and protein homology **combined**. Above Figure 3 a table with the score of these prediction are given from the top 10 proteins. The top hits were highly associated with high gene neighborhood scores and also encoded for **adhesion proteins** [6].

IN SILICO

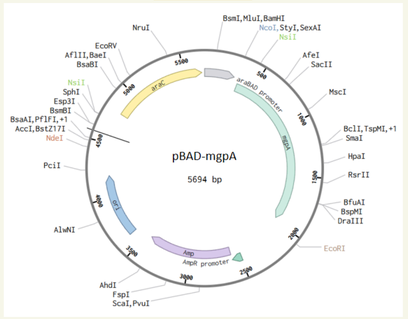


Figure 4: In silico design of the overexpression vector for the mgpA gene. Primers with specific restriction sites were used to isolate the main gene encoding for Adhesin P1, *mgpA*, from the *M. pneumoniae* genome. The **NcoI** and **EcoRI** restriction enzymes were used to introduce sticky in both the vector backbone and GOI in order to create the recombinant **pBAD-mgpA** plasmid. This construct can be transformed into *E.coli* cells, where Adhesin can be expressed.

MECHANISM OF ADHESION

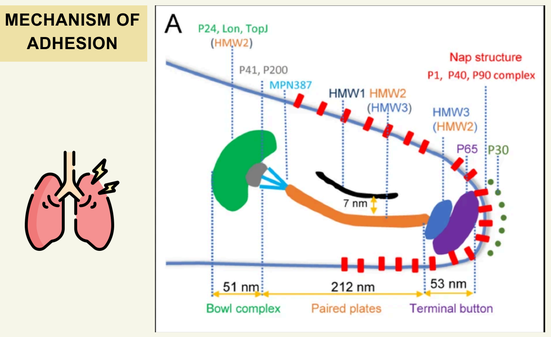


Figure 5: Attachment organelle in *M. pneumoniae* [7]. The C-terminal of P1 plays a crucial role in the attachment process. It is mostly clustered at the tip of the attachment organelle, where **P1, attached to P40/90 complex**, recognises and attaches to **sialylated oligosaccharides**. This regulates the P40/90 binding site, making P1 essential for gliding motility and adhesion [7]. **Proteins A, B, and C** act as **scaffolds,** ensuring the correct localisation of P1 in the mycoplasma membrane. Also, the proteins **HMW1, 2, and 3** are critical for the structural integrity of the entire attachment organelle. Regarding the motility in *M. pneumoniae*, it is **ATP-dependent** as well as **unidirectional** thanks to the Nap machinery [7, 8]. Moreover, the energy requirements of *mycoplasma* are directly related to its overall ability to attach to different surfaces via the relationship between the **phosphorylation rate of P1 and HMWs and glucose content** [8].

INFECTION AND IMMUNE RESPONSE

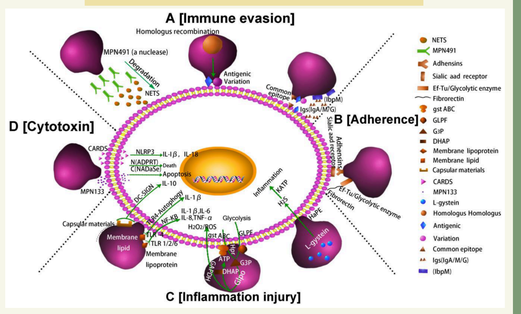


Figure 6: Mechanism of *M. pneumoniae* infection and immune response [9]. Different proteases and proteins from *mycoplasma* could dampen the **innate immune response** of the host cell by releasing proteases and degrading nucleophiles [7]. They can also release **ROS** and **hydrogen peroxide** which, in combination with their ability to trigger **active immune response, further deepen cell damage**. Additionally, the bacteria can also cleave antibodies [7, 9]. When it comes to P1 and P40/90, their structures are recognized by **Igs** proving their immunogenic value [5]. Moreover, P1 not only helps with host invasion, but also with the **escape of pathogens from Abs**. However, the rigidity and genetic variability of its N-end make it a bad drug target while the highly conserved, yet **flexible structure of the C-end** makes it very efficient for such purposes [5, 7].

CONCLUSION

The fragment was identified as part of *M. pneumoniae* by BLASTN. **Adhesin P1** was found in the DNA fragment by KBase. Other proteins found in the same gene neighborhood, also had cell adhesion properties. The structure of the amino acid sequence (from KBase) only **partially covers** the structure (7BWM) from blastp, but they do align. There was found that the KBase sequence only contains a shorter C-terminal domain of adhesin P1 and no N-terminal domain. In addition, its cooperation with P40/90 is essential for inducing adhesion and gliding motility.
Adhesin P1 was overexpressed in silico in the plasmid pBAD-sfGFP. This enables further study of the protein structure and functionality, host interactions, immune recognition and potential vaccine production. Since P1 exhibits such a central role in adhesion and inhibition of the immune response, targeting its C-terminal is a leading and sensible strategy for development of **mRNA vaccines.**

REFERENCES

Unraveling Chlamydia abortus:
Targeting the SRP-Dependent Pathway to Neutralize Pathogenicity

Noemi Ricci S5463157, Leyenne Schairer S5987164, Aurelija Kasputyte S5879434, Laura Romero S4666534,
Bart Oude Avenhuis S3773175, Adam Sundell S6511627

References

Abstract

This case study focuses on a DNA fragment set from an unknown bacterium, which was identified as belonging to the *Chlamydia abortus* species through an ***nBLAST search***.
Using theoretical, genomic, and proteomic analyses, the genes present, the proteins they encode, and their roles in pathogenicity, protein interactions, and cellular functions were examined, as well as strategies for managing infections. These analyses were performed using tools such as ***NCBI, KBase, AlphaFold and Benchling.***

Introduction

The *"Tree of Life"* is a fundamental concept in Biology which divides all life into three major groups, one of which is Bacteria. Bacteria are highly abundant, constantly evolving, and remain a major focus of study.

In this case study, a DNA fragment from an unknown bacterium was identified as ***Chlamydia abortus*** and analyzed using theoretical and bioinformatics tools, focusing on genes and proteins involved in pathogenicity.

Chlamydia abortus is a non-motile, obligate intracellular, Gram-negative bacterium that primarily infects small ruminants but can also infect pregnant women, making it a zoonotic pathogen (1).
Transmission occurs mainly through ***ingestion or aerosol inhalation*** of contaminated abortion material (2).

It has a ***biphasic life cycle***, alternating between the infectious elementary body (EB) and the replicative reticulate body (RB) (3). Several EB surface proteins, including ***polymorphic membrane proteins (Pmps)***, function as autotransporters essential for host adhesion (4).

Their correct localization depends on ***the signal recognition particle (SRP) pathway***, with ffh gene encoding the SRP protein responsible for directing membrane proteins to their proper location (5, 6).

The potential neutralization of pathogenicity was explored through ***gene overexpression*** using an in-silico design.

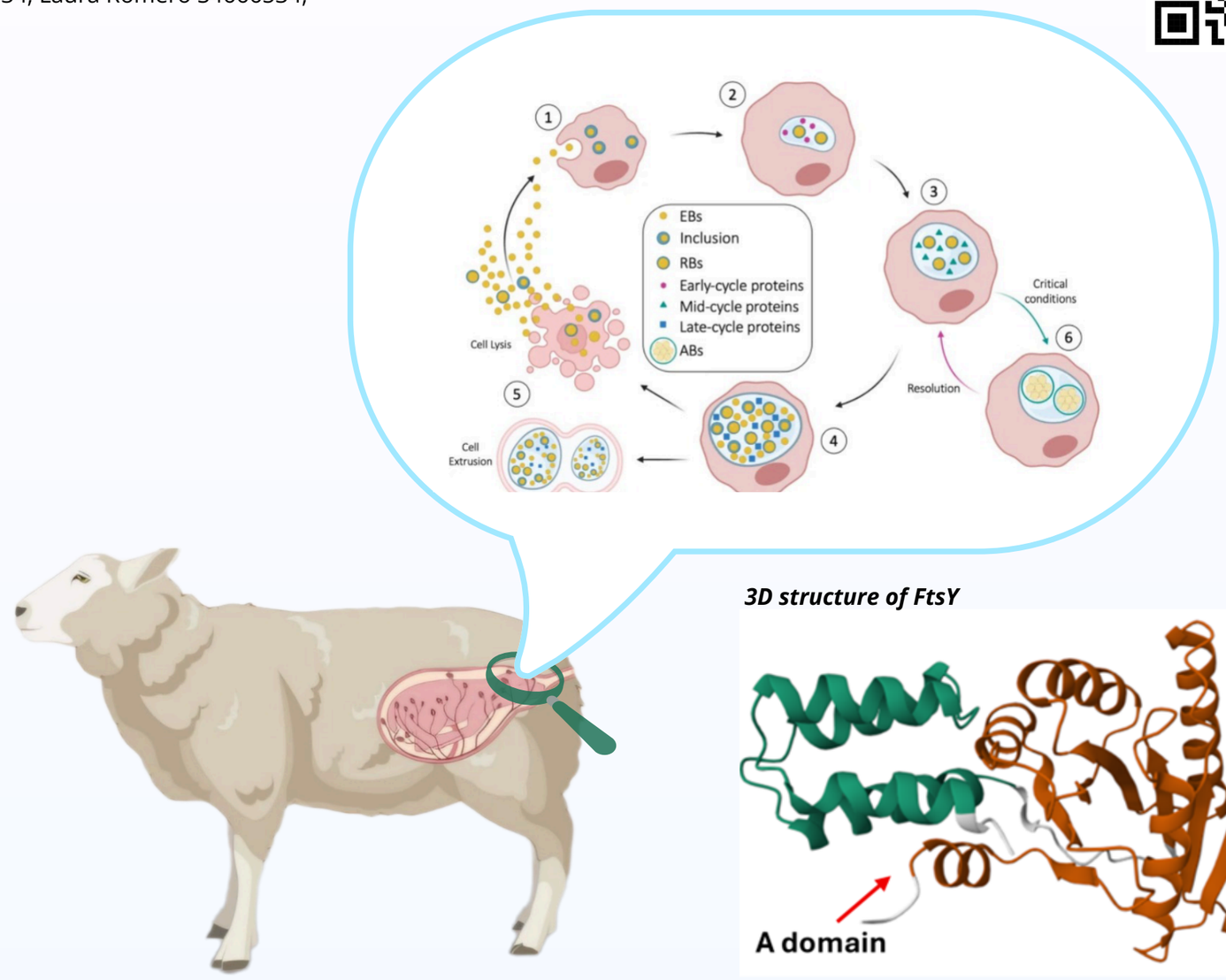


Figure 1: Cartoon representation of a pregnant sheep showing the biphasic developmental cycle of Chlamydia abortus, illustrating the attachment and entry of elementary bodies (EBs) into susceptible host cells, their conversion to reticulate bodies (RBs), and subsequent release as EBs (6).

Methods and materials

Organism and gene identification
NCBI nucleotide blast was used to identify from which organism DNA fragment set nine was from (see table 1).
Standard database (nr, etc.) (core_nt) was used and optimisation of highly similar sequences (megablast).
KBase was used for efficient and comprehensive gene annotations (see table 2) by Annotate Assembly and Re-annotate Genomes with Prokka - v1.14.5 Type: KBaseGenomes.Genome-11.1 Version: 1.

Literature research on pathogenicity
To understand the transmission and infection mechanisms of Chlamydia abortus, a ***literature review*** was conducted focusing on its origin, host targets, modes of infection, and pathogenicity, with particular attention to the SRP pathway and the roles of ffh, ftsY, and Ftsy.

Structure visualization of FtsY
The protein structure of the signal recognition particle receptor (FtsY) was found and visualized in ***AlphaFold*** (EMBL-EBI, Alphafold PSD). The three domains were highlighted with the "domain" feature in cartoon presentation mode after which the A domain was further highlighted.

In silico design of the ftsY expression vector
The nucleotide sequence of ftsY was found on ***NCBI*** and imported in ***Benchling***. ***ThermoFisher*** was used to find the sequence for the pBAD/His vector and Primers were designed in ***NCBI Primer-BLAST***.
Digestion and ligation were performed and analysis of recombinant DNA was done to check that final construct is correct.

Results

Organism identification

Scientific Name	Total Score	Query Cover	E value	Per. Ident	Accession
Chlamydia abortus S26/3	36934	100%	0	100%	CR848038.1
Chlamydia abortus	36928	100%	0	100%	CP024084.1
Chlamydia abortus	36767	100%	0	9.986%	LN554883.1
Chlamydia abortus	36767	100%	0	9.986%	CP070224.1
Chlamydia abortus	36767	100%	0	9.986%	LN554882.1
Chlamydia abortus	36767	100%	0	9.986%	CP021996.1
Chlamydia abortus LLG	36322	100%	0	9.946%	CP018296.1
Chlamydia abortus	36265	100%	0	9.941%	CP158101.1

Table 1: nBLAST results for organism identification showing the scientific name, total score, query cover, E-value, percentage identity, and accession number. The total score reflects alignment strength, query cover shows sequence overlap, the E-value indicates match significance, and percentage identity measures sequence similarity. The table identifies bacterium Chlamydia abortus as the organism fragment set 9 belongs to.

KBase genome annotation

Fragment num.	Type	Start position	Length (nt)	Strand (-/+)	Gene synonym	Full name of product
fragment_2	gene	4764	5390	-	atpD	V-type ATP synthase subunit D
fragment_2	gene	9313	9939	-	atpE	V-type proton ATPase subunit E
fragment_1	gene	6324	7190	+	COQ2	4-hydroxybenzoate polyprenyltransferase, mitochondrial
fragment_1	gene	8234	9334	+	ecdB	putative UbiX-like flavin prenyltransferase
fragment_3	**CDS**	**19390**	**19995**	**-**	**ffh**	**Signal recognition particle protein**
fragment_3	**gene**	**19390**	**19995**	**-**	**ffh**	**Signal recognition particle protein**

Table 2: Table summary of 6 Kbase Prokka genome annotation results from all 3 fragments of Fragment_set_09 with hypothetical proteins and regions containing no product removed. The CDS and gene of ffh (with product signal recognition particle protein) highlighted, as this gene and its protein is used for further analysis.

Biological process of SRP-dependent pathway

1. ***The signal recognition particle SRP*** is a ribonucleoprotein, consisting of Ffh-protein and 4.5S RNA, binds to signal sequences of translating ***virulence proteins (Pmps)*** emerging out of ribosomes.
2. RNC-SRP/FtsY GTPases ***dimerization*** at the NG domains
3. Pmps are passed through the translocation pore on the inner plasma membrane of the bacteria by the ***hydrolysis of GTP***.
4. ***Disassociation*** of the RNC-SRP/FtsY complex (7).

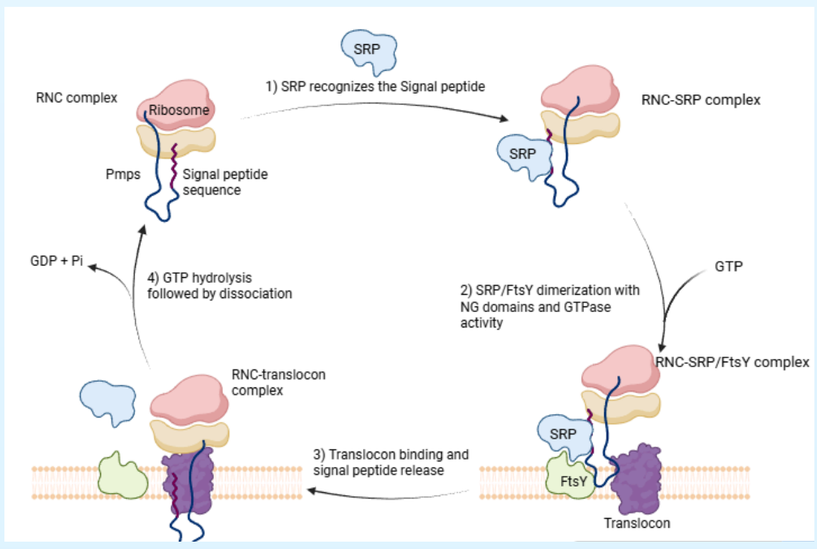


Figure 2. The four steps of the signal recognition particle (SRP) dependent pathway. The pathway starts with recognition of the signal peptide sequence on the Pmp by SRP, SRP/FtsY dimerization, translocon binding and dissociation of the SRP/FtsY complex. Created with biorender.

3D structure of FtsY

Figure 3: AlphaFold protein structure of the FtsY receptor is divided into three domains: the NG domains (N = orange, G = green) and the A domain (highlighted in the picture). The NG domain is divided into a nucleotide binding domain (N) and a GTPase domain (G) that contains nucleoside triphosphate hydrolases whereas the A-domain anchors the protein to the membrane.

In Silico Design overexpression ftsY gene

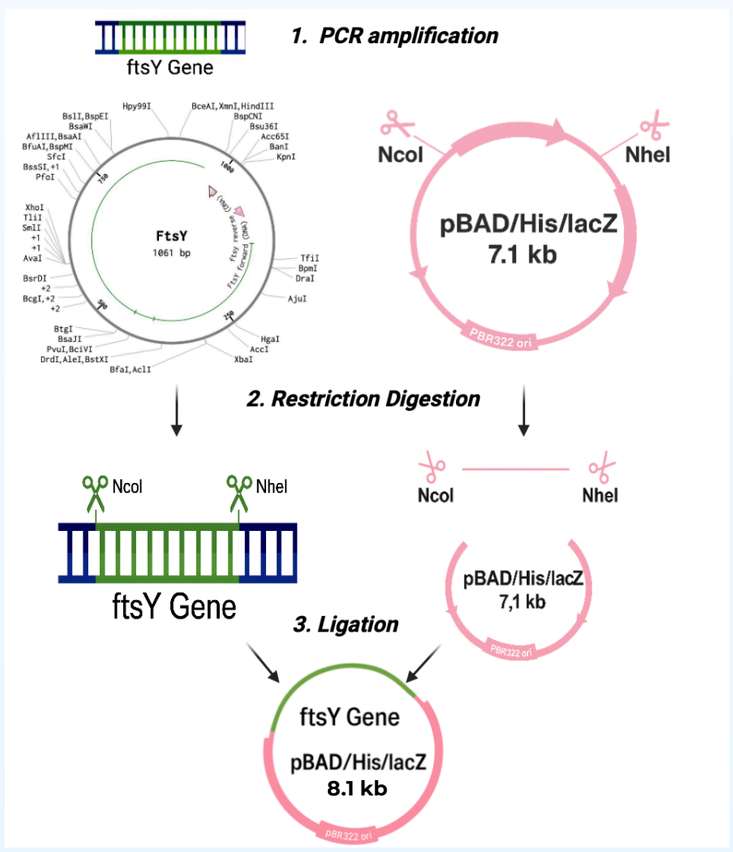


Figure 4: in Silico Design for cloning FtsY gene in pBAD vector.
It shows pBAD vector used to clone FtsY receptor gene from Chlamydia abortus S26/3 strain using NcoI and NheI restriction enzymes. The recombinant plasmid has a length of 8083 bp.

Conclusion

Nucleotide BLAST analysis of fragment set 9 showed the sequence belonged Chlamydia abortus. The ffh gene from 57 KBase annotatated genes was selected for further study because it encodes a signal recognition particle protein (SRP) involved in the correct localization of polymorphic membrane proteins (Pmps) that play a role in the ***organism's pathogenicity.***

The protein structure of the SRP receptor FtsY showed that FtsY is anchored to the inner membrane with its A domain and that the ***SRP/FtsY complex*** is made possible by the NG domains dimerizing.

The overexpression of the FtsY receptor will stimulate premature binding to the Ffh-SRP preventing the ***translocation of virulence protein Pmps.***

Further optimization of the outlined mechanism could consider post translational modification by **cleavage of N-terminus** at the A domain. This would lead to increased solubility of the FtsY receptor after overexpression ensuring selective accumulation of`FtsY in the cytosol, promoting premature binding of Ffh-SRP (8).

Ultimately, regulation of FtsY-Ffh interactions could enable **new antimicrobial therapeutic treatments** against Chlamydia Abortus infections in affected organisms.

Listeria Monocytogenes

Bioinformatic Characterization of the Genome and Proteome

Scholtens N.B. (S6125395), Rorije E. (S6050824), Wijnands S.P. (S6110533), Vermue N.C. (S5962064), Westerhof J.A.M. (S5552729), Venema R. (S5559014)
WBBY002-05.2025-2026.1 Bioinformatics



Abstract

This case study explores a DNA fragment, which was determined to be Listeria Monocytogenes using BLAST. This project aims to use bioinformatics tools to analyze the genome and proteome of L. monocytogenes to uncover how its genes and key proteins, such as PrfA and CcpA, contribute to its pathogenicity and global health impact.

Introduction

Listeria monocytogenes is a gram-positive, catalase-positive, facultative intracellular rod that can grow at refrigerator temperatures [29]. It is a major cause of foodborne illness, commonly linked to contaminated deli meats and unpasteurized dairy products. The bacterium survives in diverse environments such as soil, water, sewage, animals, and food processing plants, and can infect both humans and animals [13].

In humans, it causes listeriosis, which may lead to meningitis, sepsis, encephalitis, spontaneous abortion, or mild gastroenteritis, particularly in pregnant women, the elderly, and immunocompromised individuals [29]. Transmission occurs mainly through ingestion of contaminated food, responsible for 85–95% of cases, and can also occur transplacentally or during birth; nosocomial and person-to-person spread are rare [22].

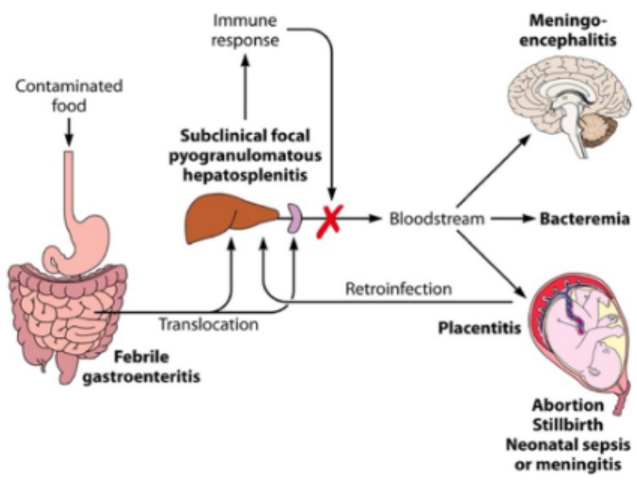


Figure 1. *Transmission of L. monocytogenes bacteria. Transmission happens via ingestion of contaminated food or transplacentally during pregnancy [23].*

Methods

NCBI

The NCBI webserver was first used to determine the organism, using a nucleotide BLAST. This was done using the standard databases and optimized for highly similar sequences.
The webserver was also used to perform an ORFfinder, using the settings: minimal ORF length 75 nt, genetic code 1. standard, and "ATG" and alternative initiation codons.

KBase

The KBase webserver was used to perform a gene annotation. This was done following the instructions of the KBase documentation guide. Prokka (Annotate Assembly and Re-annotate Genomes with Prokka – v1.14.5) was used to annotate the sequence.

Protein analysis

To analyze the protein structures, AlphaFold, InterPro, TMHMM, and the PDB webserver were used. This was done for two different genes.

In silico overexpression

The ccpA gene from L. monocytogenes was amplified using primers designed with NCBI Primer-BLAST and Benchling, incorporating restriction sites to generate sticky ends. Complementary sticky ends were introduced into the pBAD vector using the same restriction enzymes. The recombinant plasmid was transformed into Escherichia coli DH5-α . This allowed production of the CcpA protein in a bacterial system.

Results

Blast of unknown sequence

Description	Total score	Query cover	E-value	Accession number
Listeria monocytogenes strain FDA00006667 chromosome	37010	100%	0.0	CP044430.2
Listeria monocytogenes strain OB040119 chromosome	37010	100%	0.0	CP053630.1
Listeria monocytogenes strain J1817, complete genome	37010	100%	0.0	CP006599.1
Listeria monocytogenes strain OB030029 chromosome	37010	100%	0.0	CP053628.1
Listeria monocytogenes strain J1776	37010	100%	0.0	CP006598.1

Table 1: *BLAST results. The DNA fragment is located on the L. monocytogenes genome*

Gene annotation of *L.monocytogenes*

Gene	strand	start	stop	product
ahpC	+	9381	9926	Alkyl hydroperoxide reductase C
aroA	+	12556	13641	Protein AroA(G)
ccpA	+	13876	14883	Catabolite control protein A
ezrA	+	18161	19876	Septation ring formation regulator EzrA
msrC	-	17378	17866	Free methionine-R-sulfoxide reductase
murC	+	7893	9236	UDP-N-acetylmuramate--L-alanine ligase
pepS	-	9967	11079	Aminopeptidase PepS
rpsD	-	16564	17166	30S ribosomal protein S4
tyrS	+	15205	16464	Tyrosine--tRNA ligase
ysdC	+	1769	2842	Putative aminopeptidase YsdC
ytnP	+	64	840	putative quorum-quenching lactonase YtnP
ytpP	+	3168	3479	Thioredoxin-like protein YtpP

Table 2: *KBase results fragment 2*

Many proteins were found, no pathogenic proteins were found in all 3 fragments, therefore further analysis was done with ccpA found in fragment 2 and PrfA, not found in any fragment, due to its importance in pathogenicity.

Pathogenicity

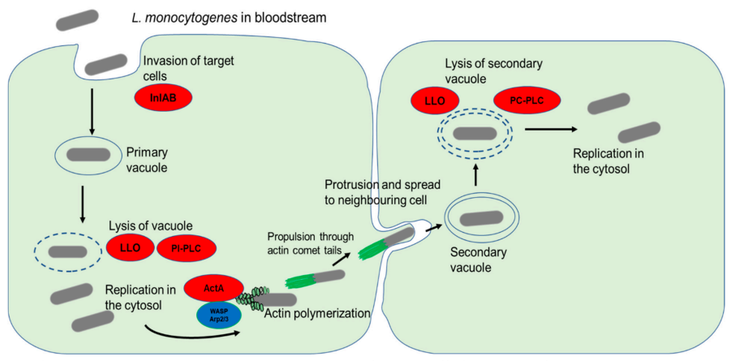


Figure 2 *L. monocytogenes invasion and intracellular spread. Surface proteins InlA and InlB enable host cell entry, after which LLO and PI-PLC lyse the primary vacuole, releasing bacteria into the cytosol for replication. ActA recruits VASP and the Arp2/3 complex to drive cell-to-cell spread, with LLO and PC-PLC lysing the secondary vacuole to infect adjacent cells [32].*

Transmission pathway

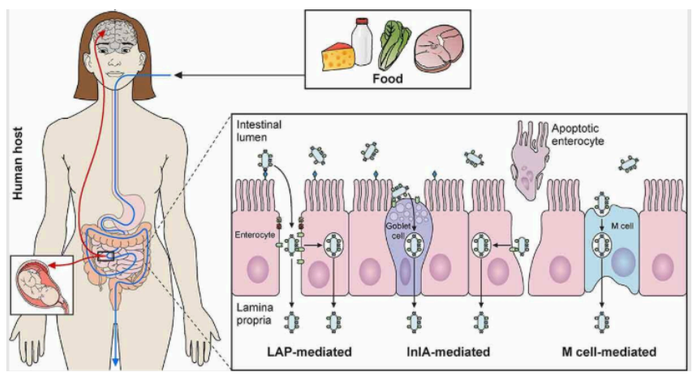


Figure 3 *The transmission pathway of Listeria monocytogenes through oral infection [12].*

Protein structures of ccpA and prfA

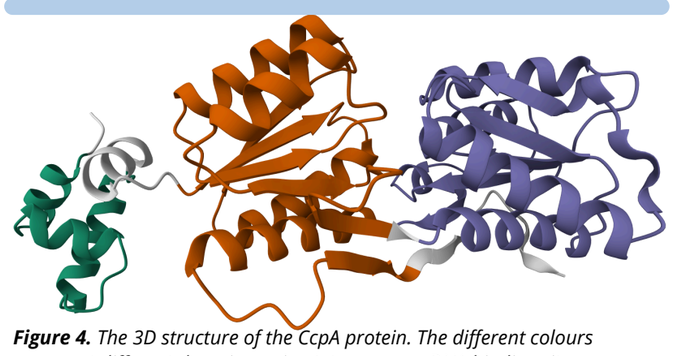


Figure 4. *The 3D structure of the CcpA protein. The different colours represent different domains, using Interpro, one DNA binding site was identified as a helix-turn-helix (green) as well as two Rossmann like folds (red, purple) which assists in cofactor binding.*

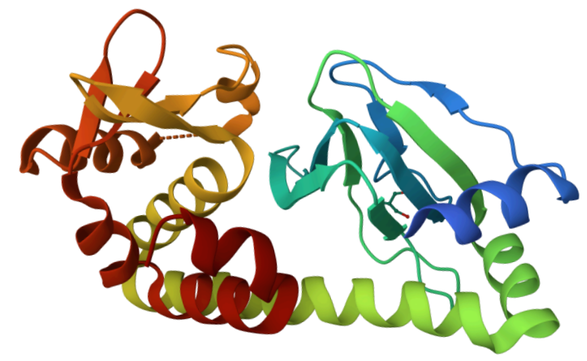


Figure 5. *The 3D structure of the PrfA protein. The colours represent the location of the N-terminal (blue) and C-terminal (red). Analysis of the protein sequence in Interpro revealed a helix-turn-helix important in DNA binding and a jelly roll fold which binds cofactors.*

In silico vector design

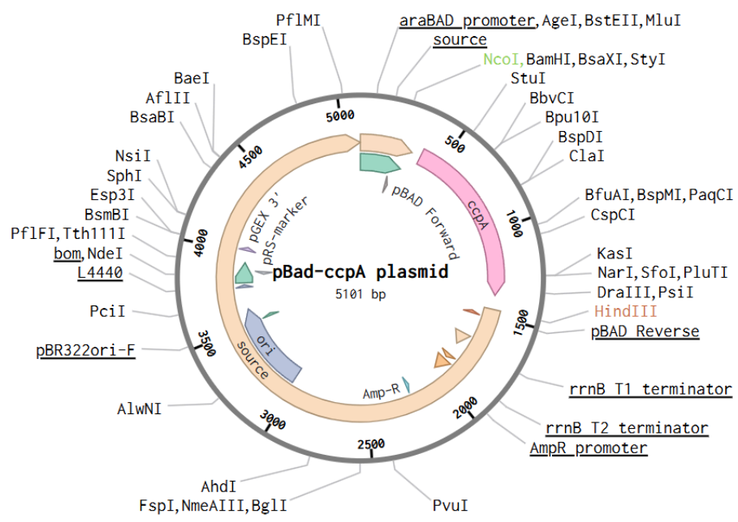


Figure 6. *The pBAD plasmid containing the ccpA gene, indicated in pink. The restriction enzymes used are highlighted in green and brown.*

Global health

Even though listeriosis is a relatively rare disease with 0.1 to 10 cases per million per year [4; 38), Listeria monocytogenes is still among the top causes of foodborne illness-related hospitalizations in North America and Europe [18].
There is no human vaccine [24], so global health strategies rely on safe food handling and avoidance of high-risk foods [9]. Treatment typically involves ampicillin, though antibiotic resistance is a growing concern [33]. Research into new therapeutic targets about PrfA inhibitors and FDA-approved drugs is ongoing, but shows promise. [19; 11; 27].

Conclusion

A detailed analysis of Listeria monocytogenes was performed using various bioinformatic tools and resources. This revealed that it is a foodborne pathogen capable of surviving and replicating at low temperatures, making it a big cause for foodborne outbreaks.

Analysis of the ccpA and prfA genes and their structures showed insights into the organism's metabolic regulation and virulence factors. Moreover, a construction of an overexpression vector for ccpA was made to study how the protein works and to explore it as a possible target for drug design. Lastly, Research about listeria monocytogenes' pathogenic mechanisms, revealed that L. monocytogenes infects via the transmission pathway.

Overall, Listeriosis' global health impact, with it specifically being risky for vulnerable groups, highlights the importance of food safety and proper therapies.

References

Legionella pneumophila, analysis of its genes and proteins

De Vries, S.P. (S5586216), Reitsma. J (S5587905), Van Os. R (S5507707), Wijnberg. N (S5616026), Zmpainou. C (S4615077). Bioinformatics, WBBY002

Abstract

In this case study an unknown DNA fragment is traced back to the species *Legionella pneumophila* using BLAST. The content of the DNA fragment and its significance is studied using an array of online tools and a literature study.

Introduction

In this case study, an unknown DNA fragment, originating from the pathogen Legionella pneumophila, was analyzed using bioinformatics tools to gain detailed insights into its genes and corresponding proteins.

Legionella pneumophila is an aquatic bacterium that survives within macrophages and amoebae by inhibiting the fusion of the phagosome with the lysosome. (Vogel & Isberg, 1999) Within host cells, it establishes a replicative vacuole surrounded by rough endoplasmic reticulum (ER), which serves as the site of bacterial multiplication. (Vogel & Isberg, 1999) The genome, approximately 3.9 Mb in size, encodes multiple secretion systems and numerous virulence factors. (Steinert et al., 2002) Pathogenicity is multifactorial, with an estimated 10% of the genome contributing to virulence. (Ott, 1994) Human infection occurs through the inhalation of contaminated aerosols, leading to Legionnaires' disease (Vogel & Isberg, 1999), characterized by intracellular replication of the bacteria within vacuoles that evade lysosomal fusion (Steinert et al., 2002).

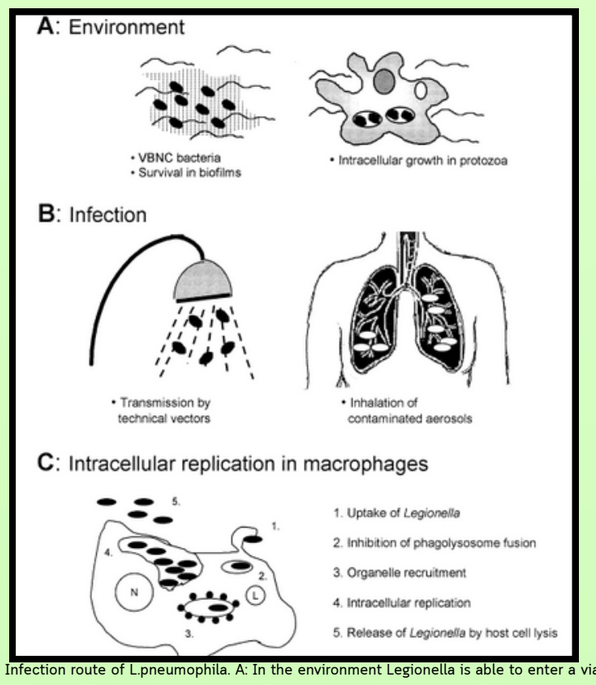


Figure 1. Infection route of L.pneumophila. A: In the environment Legionella is able to enter a viable but non-culturable (VBNC) state. The bacteria persist in biofilms and grow within protozoa. B: Upon transmission by technical vectors (showers, air conditioning systems, cooling towers, etc.) Legionella colonizes the human respiratory tract. C: After uptake by macrophages Legionella replicates within a maturation-blocked vacuole. Finally the bacteria are released by host cell lysis. Abbreviations: N, nucleus; L, lysosome (Steinert et al., 2002).

Materials and methods

BLASTn

A BLASTn was conducted on the unknown DNA fragment using standard databases and optimizing for high similarity.

kBase & ORF-finder

After blasting, Kbase was performed to identify genes in the DNA fragment. These genes were further studied using ORF-finder with the following settings: 300nt minimal ORF length, start codon: ATG and alternative.

Literature study

A literature study was conducted to better understand *L. pneumophilia,* its pathogenicity and the selected ORF.

In silico overexpression

To design an overexpression vector, benchling, an online tool, was used. To isolate our gene from the fragment, primers were designed containing restriction enzymes, which ligate the gene in the right orientation to our vector (pRSET/C). This complete vector is later intergrated inBL21(DE3) E. col to produce the protein.

3D-structure of ABC transporter ATP-binding protein uup

The 3d structure was gained from PyMOL in Alphafold

Results

BLASTn of the unknown sequence

Name & strain	Total score	Query cover	E-value	Accession number
Legionella pneumophila strain Philadelphia_4, complete genome	36934	100%	0.0	CP015931.1
Legionella pneumophila subsp. pneumophila strain Lp02 dksA1 chromosome, complete genome	36934	100%	0.0	CP184461.1
Legionella pneumophila subsp. pneumophila ATCC 43290, complete genome	36934	100%	0.0	CP003192.1
Legionella pneumophila strain SPF544 chromosome, complete genome	36934	100%	0.0	CP174245.1
Legionella pneumophila strain lpg1386 chromosome, complete genome	36934	100%	0.0	CP169561.1
Legionella pneumophila strain AUSMDU00010536 isolate SBT211 chromosome, complete genome	36934	100%	0.0	CP045974.1
Legionella pneumophila strain Philadelphia_2, complete genome	36934	100%	0.0	CP015929.1
Legionella pneumophila strain E9_O, complete genome	36934	100%	0.0	CP015956.1
Legionella pneumophila strain C4_S, complete genome	36934	100%	0.0	CP015935.1

Table 1. BLASTn results, the DNA fragment is located on the genome of *L. pneumophila*

Gene annotation of *L. pneumophila* fragment

Type	Function	Start	Strand	Length
gene	hypothetical protein	337	-	315
gene	Histidinol-phosphate aminotransferase	497	+	1110
gene	Putative pterin-4-alpha-carbinolamine dehydratase	1643	+	342
gene	Protein translocase subunit SecF	3074	-	918
gene	Protein translocase subunit SecD	4943	-	1857
gene	Sec translocon accessory complex subunit YajC	5300	-	336
gene	Queuine tRNA-ribosyltransferase catalytic subunit 1	6279	-	834
gene	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	7339	-	1014
gene	hypothetical protein	7869	-	414
gene	ABC transporter ATP-binding protein uup	9804	-	1854
gene	hypothetical protein	10356	-	492
gene	2-iminobutanoate/2-iminopropanoate deaminase	10745	-	387
gene	Bifunctional (p)ppGpp synthase/hydrolase SpoT	12892	-	2148
gene	DNA-directed RNA polymerase subunit omega	13239	-	204
gene	Guanylate kinase	13919	-	630
gene	hypothetical protein	14791	-	867
gene	Ribonuclease PH	15092	+	708
gene	Twitching mobility protein	17163	-	1035
gene	Pyridoxal phosphate homeostasis protein	17327	+	696
gene	Pyrroline-5-carboxylate reductase	18019	+	789
gene	hypothetical protein	19152	+	576

Table 2. kBase and ORF-finder results to determine the genes. Different genes were found, including the gene that expresses ABC transporter ATP-binding protein uup

Pathogenicity pathway of *L. pneumophila*

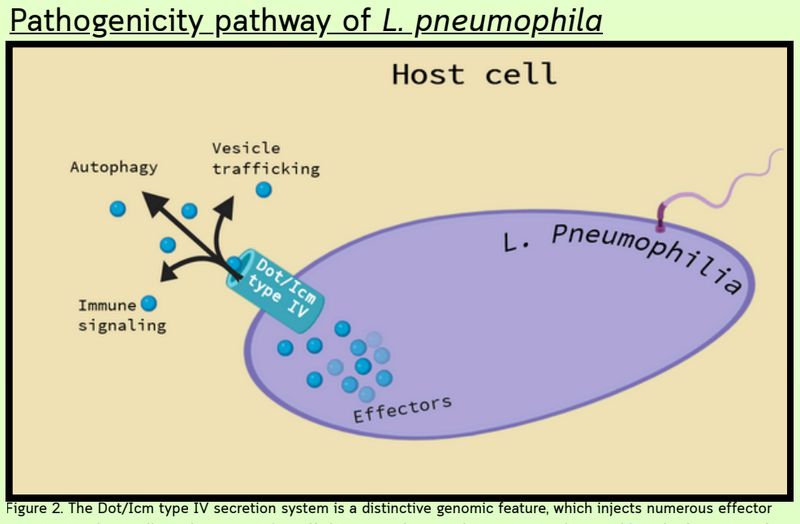


Figure 2. The Dot/Icm type IV secretion system is a distinctive genomic feature, which injects numerous effector proteins into host cells to disrupt vesicle trafficking, autophagy, and immune signaling, enabling the formation of a replicative niche within the Legionella-containing vacuole (LCV)(Romanov & O'Connor, 2024). Created with BioRender.

In silico overexpression of ABC transporter ATP-binding protein uup

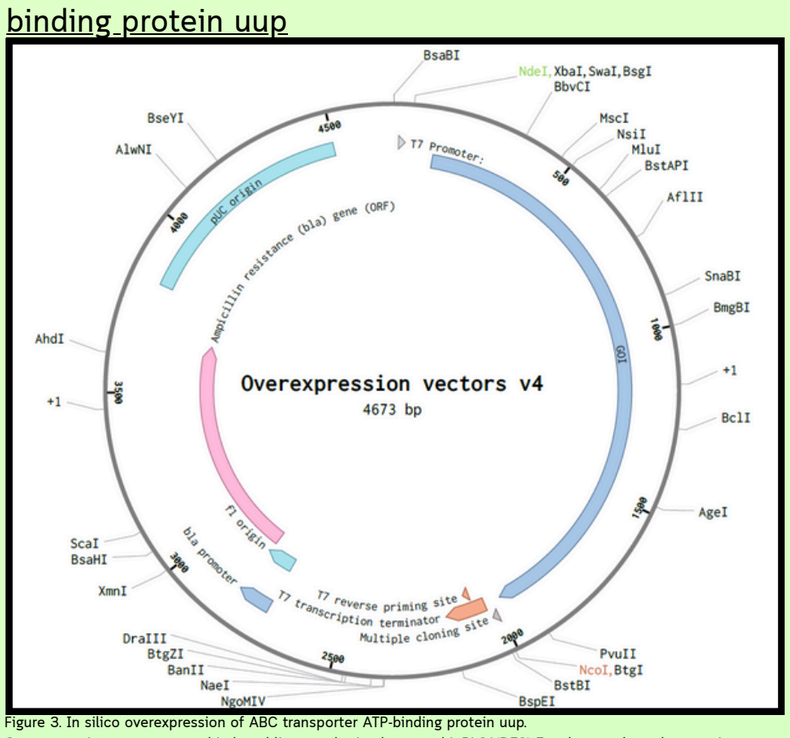


Figure 3. In silico overexpression of ABC transporter ATP-binding protein uup.
Overexpression vector created in benchling, can be implemented inBL21(DE3) E. col to produce the protein.

3D-structure of ABC transporter ATP-binding protein uup

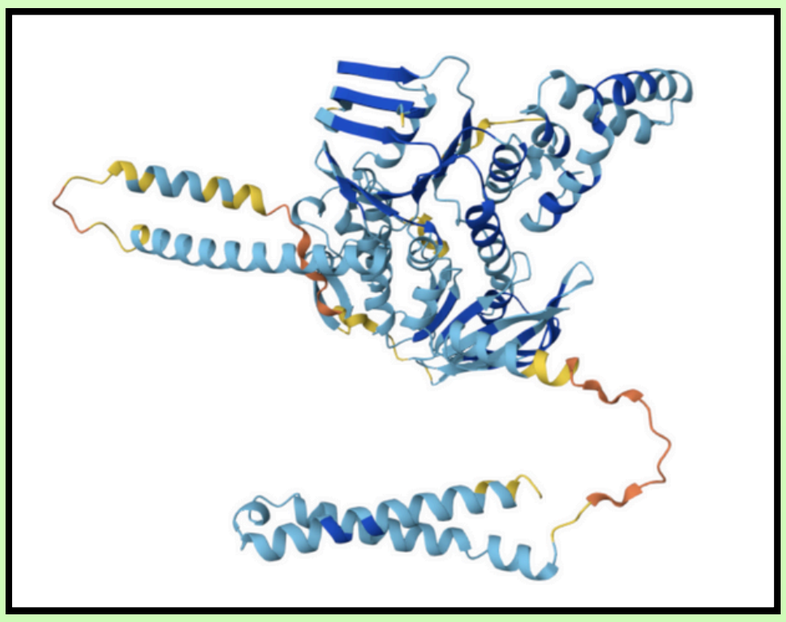


Figure 4. the constructed three dimensional stucture in AlphaFold showing a head and tail structure. It contains 2 ATP binding domains and a tRNA binding domain.

Conclusions

The unknown DNA fragment was blasted to trace it back to organism *Legionella pneumophila.* Kbase and ORF-finder were used as bioinformatics tools to find genes and proteins encoded by the fragment.

Literature studies found that the ABC transporter ATP binding protein uup probably helps maintaing genomic stability and supporting stress responses. The Dot/Icm type IV secretion system is responsible for the pathogenic effect by injecting effectors into the host cell.

Overexpression of the selected protein, the ABC transporter ATP-binding protein Uup, is relevant for both research and potential antibacterial applications. By overexpressing this protein, its specific function and ways to inhibit its activity can be researched. Since Uup is involved in DNA repair processes, inhibiting it could disrupt genomic stability and reduce bacterial viability, making it a potential target for antibacterial strategies.

university of groningen

References

Fields, B. S., Benson, R. F., & Besser, R. E. (2002). Legionella and Legionnaires' disease: 25 years of investigation. *Clinical Microbiology Reviews, 15*(3), 506-526. https://doi.org/10.1128/CMR.15.3.506-526.2002

Gomez-Valero, L., Rusniok, C., Jarraud, S., Vacherie, B., Rouy, Z., Barbe, V., Medigue, C., Etienne, J & Buchrieser, C. (2011). Extensive recombination events and horizontal gene transfer shaped the Legionella pneumophila genomes. *BMC Genomics, 12*, 536. https://doi.org/10.1186/1471-2164-12-536

Graham, C.I., MacMartin, T.L., de Kievit, T.R., Brassinga, A.K.C. (2023). Molecular regulation of virulence in Legionella pneumophila. *Molecular microbiology*, 121(2), 167-195. https://doi.org/10.1111/mmi.15172

Molecular & Cellular Proteomics. (2022). New global insights on the regulation of the biphasic life cycle and ClpP-dependent proteolysis in Legionella pneumophila. https://www.mcponline.org/article/S1535-9476%2822%2900041-X/fulltext

Ott, M. (1994). Genetic approaches to study Legionella pneumophila pathogenicity. *FEMS microbiology reviews*, 14(2), 161-176. https://doi.org/10.1111/j.1574-6976.1994.tb00085.x

Palusińska-Szysz, M., & Simões, M. (2024). Legionella pneumophila–Transmission, pathogenesis, host-pathogen interaction, prevention and treatment. *Frontiers in Microbiology,* 15, Article 1364620. https://doi.org/10.3389/fmicb.2024.1364620

Proteopedia. (n.d.). ABC transporter. Retrieved from https://proteopedia.org/wiki/index.php/ABC_transporter

Romanov, K. A. & O'Connor, T. J. (2024). Legionella pneumophila: A Rosetta stone to understanding bacterial pathogenesis. Journal of Bacteriology, 206(12). https://doi.org/10.1128/jb.00324-24

Sahr, T., Rusniok, C., Impens, F., Oliva, G., Sismeiro, O., Coppée, J. Y. & Buchrieser, C. (2017). The Legionella pneumophila genome evolved to accommodate multiple regulatory mechanisms controlled by the CsrA system. PLoS Genetics, 13(2), e1006629. https://doi.org/10.1371/journal.pgen.1006629

Schneider, E., & Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiology Reviews,* 22(1), 1-20. https://doi.org/10.1111/j.1574-6976.1998.tb00358.x

Steinert, M., Hentschel, U., & Hacker, J. (2002). Legionella pneumophila: an aquatic microbe goes astray. *FEMS microbiology reviews*, 26(2), 149-162. https://doi.org/10.1111/j.1574-6976.2002.tb00607.x

Tiaden, A., Spirig, T., Weber, S. S., Brüggemann, H., Bosshard, R., Buchrieser, C. & Hilbi, H. (2021). A horizontally acquired Legionella genomic island encoding a LuxR-type regulator links quorum sensing and virulence. Molecular Microbiology, 115(4), 678-695. https://doi.org/10.1111/mmi.14770

Vogel, J.P., Isberg, R.R. (1999). Cell biology of Legionella pneumophila. *Current opinion in microbiology*, 2(1), 30-34. https://doi.org/10.1016/s1369-5274(99)80005-8

Helena Kiefer Diaz (S5948363)
Clara Serrano Bach (S5845734)
Eleni Antonopoulou (S5853052)
Eleonora Mendozza (S5947677)
Marija Mihaljević (S3893812)
Catarina Mendes (S5340667)

EXPLORING MYCOPLASMOIDES PNEUMONIAE THROUGH BIOINFORMATICS

From unknown DNA fragment to functional gene analysis

university of groningen

Abstract

This case study explores the contents of an unknown DNA fragment. BLAST analysis revealed that the sequence originates from **Mycoplasmoides pneumoniae**. Further investigation using bioinformatic tools such as **NCBI, RCSB, KBase, Benchling, InterPro, and AlphaFold** identified three key proteins: **a high-molecular-weight cytadherence protein, the chaperone protein DnaK, and the ATP-dependent 6-phosphofructokinase**. These proteins are involved in **adhesion, protein folding, and glycolysis,** respectively, highlighting the bacterium's adaptations for host dependence and survival.

Introduction

M. pneumoniae is a **wall-less, pleomorphic** bacterium from the class Mollicutes and the main cause of respiratory infections such as "**walking pneumonia**." (Waites & Talkington, 2004). It has one of the **smallest genomes** of any self-replicating bacterium (~816 kb, 687 genes) and relies heavily on its host for essential nutrients and energy due to **reductive evolution**. As it lacks a **peptidoglycan cell wall**, it maintains a **cholesterol rich membrane** obtained from the host.

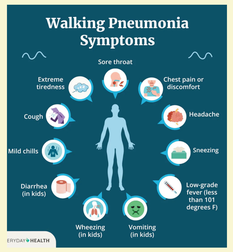


Figure 1: Walking pneumonia symptoms

The aim of this study was to **identify the origin of an unknown DNA fragment** using **bioinformatic tools** and to **analyze the biological functions** of its encoded **genes and proteins**. By examining **predicted ORFs, encoded proteins, phylogenetic relationships**, and **protein structure**, the DNA sequence was linked to functions contributing to the organism's **pathogenicity.**

Materials and Methods

Organism Identification and Gene Annotation

A BLASTx search of DNA Segment 21 was performed against the UniProt/ Swiss-Prot database to identify the organism.

To further analyze the sequence:

- ORFs were predicted for each of the three fragments within the segment, using the NCBI ORF Finder.
- From each fragment, the longest ORF was selected for further analysis using BLASTp, to identify homologous proteins.

Phylogenetic Analysis

- The ATP-dependent 6-phosphofructokinase (pfkA) sequence was compared across species using BLASTp.
- 25 homologous sequences were selected and aligned using Clustal Omega
- A phylogenetic tree was generated and visualized using iTOL to explore evolutionary relationships.

In Silico Cloning

- The pfkA gene was cloned in silico into the pET27b+R expression vector using Benchling.
- Vector sequence was retrieved from Addgene and annotated in Benchling.
- Primers containing BamHI and EcoRI restriction sites were designed using Primer-BLAST.
- Virtual PCR, restriction digestion, and ligation were carried out in Benchling's Assembly Wizard to create pET27b+R_pfkA.

Structural and Functional Protein Analysis

- The protein sequence of pfkA in M. pneumoniae was analyzed using InterProScan to gain insight into the function of the protein and its regulation.
- AlphaFold was utilized to obtain a prediction of the 3D model.

Literature research on human infection:

To better understand M. pneumoniae pathogenesis, a literature review was conducted on its interaction with human cells, focusing on adherence, toxin activity, inflammation, and immune evasion

References

Results

BLASTp Results of longest ORFs

Fragment	ORF ID	nt length	aa length	Encoded Protein	BLAST match (organism/strain)	% ident
1	ORF22	2,883	960	protein Cytadherence	*Mycoplasmoides pneumoniae* M129	99.90%
2	ORF52	1,728	576	protein Chaperone protein DnaK	*Mycoplasmoides pneumoniae* M129	100%
3	ORF29	987	329	protein Probable ATP-dependent 6-phosphofructokinase	*Mycoplasmoides pneumoniae* M129	100%

Table 1: Longest ORFs of each DNA fragment analysed with BLASTp

The longest ORFs identified correspond to key *Mycoplasmoides pneumoniae* proteins: ORF22 encodes an adhesin critical for host cell attachment, ORF52 encodes the molecular chaperone protein DnaK, involved in protein folding under stress, and ORF29 encodes the enzyme ATP-dependent 6-phosphofructokinase, a central regulator in glycolysis. The high sequence identity confirms that the DNA fragments originate from *Mycoplasmoides pneumoniae.*

Phylogenetic tree of phosphofructokinase homologs



Figure 2: Phylogenetic tree of ATP-dependent 6-phosphofructokinase (PFK) homologs for 25 bacterial species:

The PfkA sequence from Mycoplasmoides pneumoniae clusters with other Mollicutes, originating from Firmicutes. Unlike the highly regulated ATP-PFKs of Proteobacteria, M. pneumoniae's enzyme functions constitutively due to its reduced genome. The tree illustrates the evolutionary plasticity of the PFK superfamily, shaped by phosphate-donor switching, regulatory diversification, and occasional horizontal gene transfer, while maintaining a conserved role in glycolysis across bacterial lineages.

Recombinant plasmid pET27b+R_GOI (6,378 bp) *in silico*

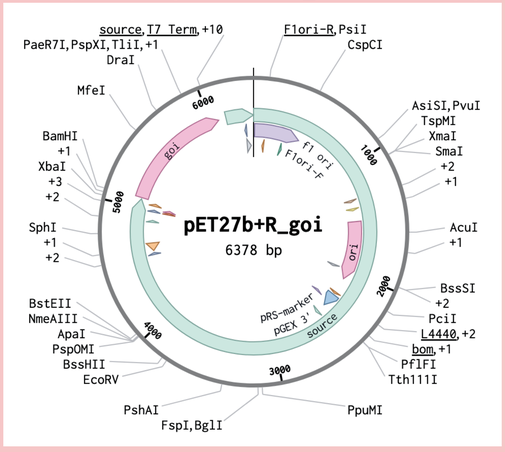


Figure 3: Benchling map of the pET27b+R_GOI construct (6,378 bp) showing the GOI insert (pink) cloned into the MCS using BamHI (5') and EcoRI (3') sites.

The recombinant plasmid pET27b+R_GOI (6,378 bp) was built by inserting the gene of interest GOI (pink) into the vector between with the BamHI and EcoRI restriction sites. The construct includes a T7-promoter and a His-tag for protein expression and purification in E.coli. The gene is inserted in the correct orientation and reading frame, confirming a successful design.

Structural and Functional Protein Analysis

Function of pfkA

- Catalyzes F6P → F1,6BP, committing F6P to glycolysis (occurs in cytoplasm)
- Binds ATP and Mg-ATP at the catalytic site.
- Switches between open and closed conformations upon ligand binding.
- Allosterically regulated:
  → Activated by ADP/AMP
  → Inhibited by ATP/citrate

A predicted 3D model was obtained, using AlphaFold, to assess the structural features of the M. pneumoniae phosphofructokinase. The probable ATP-dependent 6-phosphofructokinase protein consists of two distinct domains. Domain 1 spans residues 7-142 and 265-312, while Domain 2 covers residues 145-260. The two domains belong to two distinct CATH homologous superfamilies.

Alphafold model of phosphofructokinase domains



Figure 4: Domain organization of the probable ATP-dependent 6-phosphofructokinase protein predicted by AlphaFold. Domain 1 is green, and Domain 2 is orange.

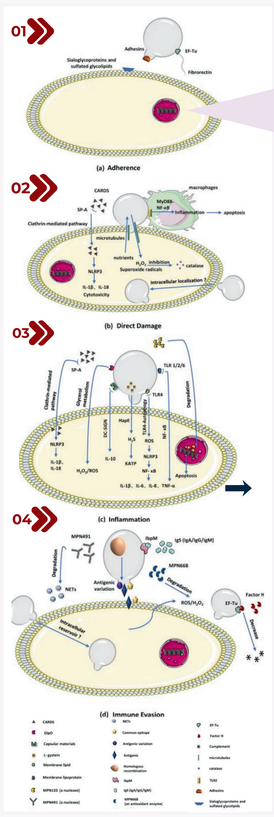
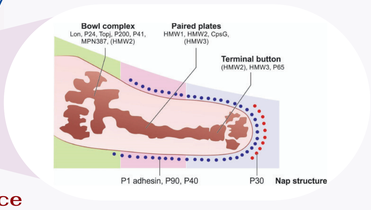


Figure 5. Pathogenic Mechanisms of Mycoplasmoides pneumoniae:

(1) Adherence via adhesins and EF-Tu, (2) Cytotoxicity by CARDS & ROS, (3) Inflammatory signaling through TLR and NLRP3 pathways (4)Immune evasion via nucleases, antioxidants, and antigenic variation.

Pathogenicity



01 Adherence

- *M. pneumonia* **adhesion causes cell damage**
- **Adhesins** tightly **bind** to sialylated sugars on our **airway cells** steal nutrients and hijack the cell's metabolism
- EF-Tu proteins can also bind to fibronectin and improve colonisation

Cytadherence = the ability of a bacterium to stick to the surface of host cells

- *M. pneumoniae* cannot move on its own ➡ **terminal organelle**
- **complex (P1, P30, P116, HMW) → mediates host attachment**
- glides slightly along surfaces

START INFECTION

- **outside** ➡ adhesins **P1, P30** are the **main sticky proteins** + accessory proteins (help P1 & P30)
- **inside** ➡ gives it **structure & stability**

terminal button +paired plates + bowl complex

02 Direct Damage

- **CARDS toxin** → triggers **inflammation** and **ciliary damage**
- **hydrogen peroxide (H_2O_2) and superoxide**

→ damage membranes, proteins, and DNA → holes → **cell injury & death**

03 Inflammation (Over-active)

- **toxins** → irritate **cell** → activate **NLRP3** → **cell releases IL-1β & IL-18**

→ **swelling, mucus, cough, tissue irritation**

04 Immune Evasion

- **Escapes neutrophil traps** (NETs) → Produces **nuclease** (MPN491) that cuts apart DNA webs, freeing the bacterium
- **Neutralizes oxidative attacks** → Makes **antioxidant enzyme** (MPN668) that breaks down peroxides (H_2O_2) and protects from ROS
- **Avoids antibody recognition** → Performs **homologous recombination**, changing its surface antigens so immune cells can't detect it easily

Conclusion

The unknown DNA fragment was identified as originating from *Mycoplasmoides pneumoniae* through BLAST analysis of the largest ORFs. *M. pneumoniae* is a small, wall-less bacterium that infects the respiratory tract using a specialized attachment organelle.

Using bioinformatic tools key proteins were identified — including a high-molecular-weight **cytadherence protein** essential for host attachment, the **molecular chaperone DnaK**, and the **ATP-dependent 6-phosphofructokinase** enzyme vital for glycolysis. Literature Research revealed pathogenicity involves adhesion-driven cell damage and virulence factors such as the CARDS toxin and ROS, which trigger inflammation and tissue injury.

Further analysis revealed evolutionary relationships of the pfkA gene, and in silico cloning of this gene into the pET27b+R vector was designed to enable functional protein expression. Future research may focus on overexpressing and characterizing this enzyme as a potential metabolic drug target, alongside exploring surface adhesins for clinical treatments of "walking pneumonia".

Pathogenicity, proteins and other characteristics of Campylobacter jejuni

rijksuniversiteit groningen

Veenema, L.C. (S5828740), Speek, I.L.M. (S6107346), Koopmans, C.E. (S5845653), Jorritsma, A.M. (S5623332), Andringa, I.R. (S5999502), Jakubauskaite, G. (S5203376)

Abstract

This case study will focus on identifying and analysing an unknown DNA sequence. For this research different methods of bioinformatics are used like BLAST, KBase and In silico design. The analysed sequence showed the identification of the bacterium species Campylobactus jejuni (C.jejuni).Further research into the proteins the sequence encodes was performed in order to connect the organism to pathogenicity and the consequences for humans.

Introduction

In this case study, we conducted a comprehensive bioinformatic analysis of an unknown DNA fragment. We determined that the sequence originates from *Campylobacter jejuni* and codes for a *hemagglutination activity domain protein*. C. jejuni is a pathogenic gram negative bacteria that is one of the most common causes of food poisoning. It is estimated that there are 96 million Campylobacter infection cases annually worldwide (Hameed, 2019). The food poisoning caused by C. jejuni is rarely life-threatening, except susceptible individuals. Infection with C. jenjuni can cause enteritis, or inflammation of the small intestine. We used various bioinformatic tools to find the protein encoded by the DNA fragment and to clarify its role in the symptoms of food poisoning caused by C. jejuni.

Methods

BLAST and ORF-finder

A nucleotide BLAST was performed using NCBI BLAST on the fragments from DNA fragment set 8. To further identify the genes present and subsequent proteins, an ORF-finder was used on fragments 1 and 2 and the two biggest ORFs discovered from each fragment were used to perform an BLASTp in order to discover the proteins they encode.

Literature research on the organism and its pathogenicity

To understand more about Campylobacter jejuni, a literature study was conducted on the origin
of the bacterium and its pathogenic function. The hemagglutination activity domain protein was further investigated within this bacterium and in other prokaryotic organisms as well.

In silico overexpression

The CDS of the target gene from Campylobacter jejuni subsp. jejuni strain 35925 (CP010906, positions 691230–694553) was identified using NCBI. Primer design included 500 bp flanking regions, optimized for 40–60% GC content. Primers were generated via NCBI Primer-BLAST using parameters: PCR size 3–4 kb, annealing temperature 55–65 °C. BamHI and EcoRI restriction sites (GGATCC, GAATTC) with upstream CACACA motifs were added to the forward and reverse primers, respectively. The gene was cloned into the pRSET-A vector containing a 6×His tag, MCS, and T7 promoter, followed by insertion via restriction digestion and assembly.

Domain structure visualization

The structure of the FahB domain of the hemagglutination activity domain protein was visualized by using the Protein Data Bank and Alphafold. For this the sequence from Kbase was used and the structure obtained was compared to the FhaB structure in Bordetella sp.02P26C-1

Results

BLASTn of DNA fragment set 8

Scientific name	Total score	Query cover	E value	Accession number
Campylobacter jejuni subsp. jejuni strain 35925	36934	100%	0.0	CP010906.1
Campylobacter jejuni strain NCTC 12661	36934	100%	0.0	CP028911.1
Campylobacter jejuni subsp. jejuni strain ATTC 35925	36934	100%	0.0	CP020045.1

Table 1: BLASTn results obtained through NCBI-BLAST. A BLASTn was performed on NCBI, with all the fragments of set 8 as input. The table summarizes the significant results from this BLAST, which indicate that the DNA fragments originate from Campylobacter jejuni.

BLASTp of ORF26

Cluster representative sequence	Total score	Per. identity	E value	Accession number
Hemagglutination activity domain protein [Campylobacter jejuni subsp. jejuni]	2205	100%	0.0	AJP35212.1
Filamentous hemagglutinin N-terminal domain-containing protein [Campylobacter jejuni]	1828	84.43%	0.0	EDO74280080.1
Filamentous hemagglutinin N-terminal domain-containing protein[Campylobacter jejuni]	1787	83.11%	0.0	EAK0933764.1

Table 1: BLASTp results of ORF26 from fragment 2, obtained through NCBI-BLAST. Hemagglutinin activity domain protein from C. jejuni shows the highest percentage identity, as well as a query cover of 100%. The ORF likely encodes a protein with hemagglutination activity, but which one it is exactly, remains unclear based on these results.

Pathogenicity of Campylobacter jejuni

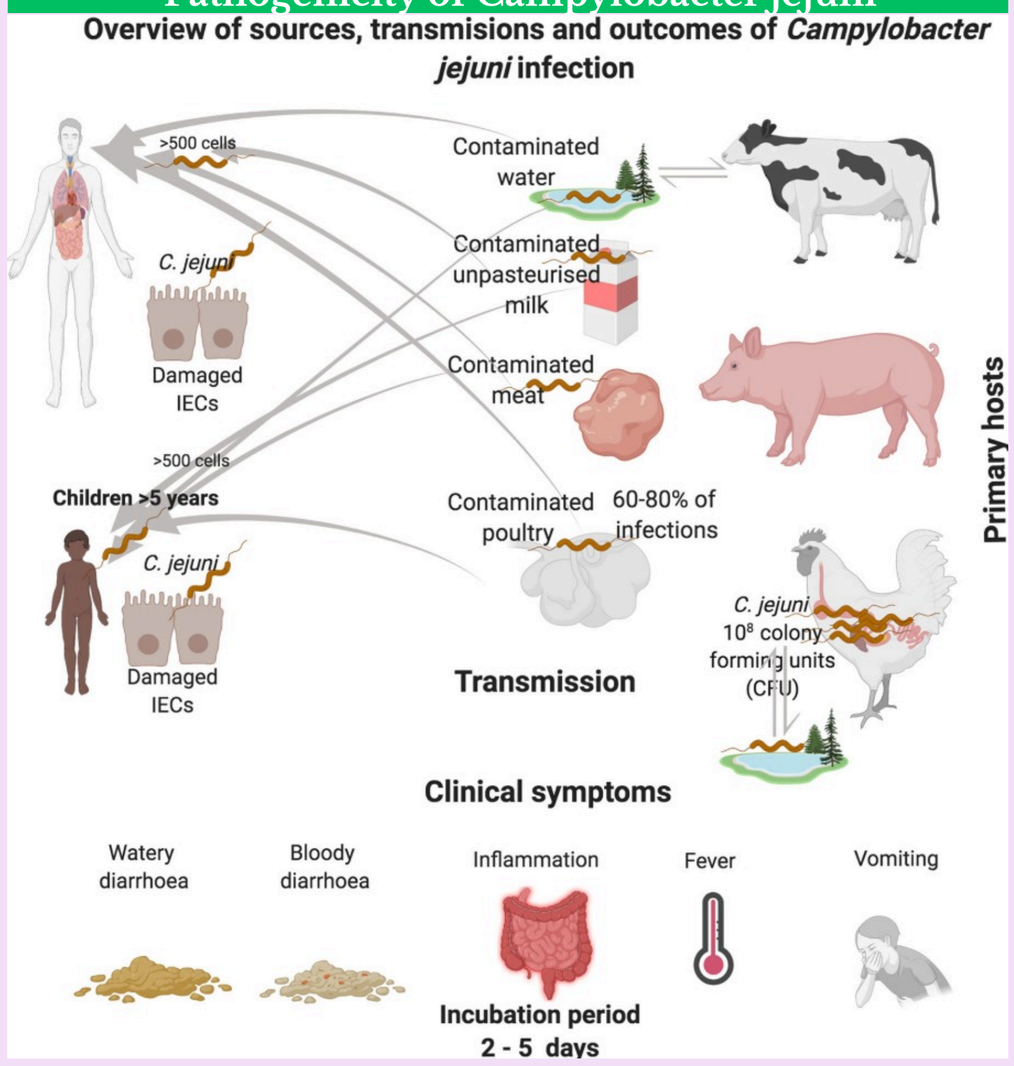


Figure 1: Summary of the Campylobacter jejuni infection in humans. This figure showcases the way in which C. jejuni is able to contaminate humans.

FhaB domain structure

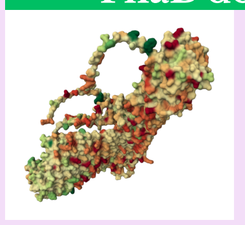


Figure 2A: 3D structure found in Alphafold of Filamentous hemagglutinin N-terminal domain-containing protein of Bordetella sp.02P26C-1. The colouring was set for hydrophobicity; green for hydrophobic (non-polar) amino acids, red for charged (hydrophilic) amino acids.

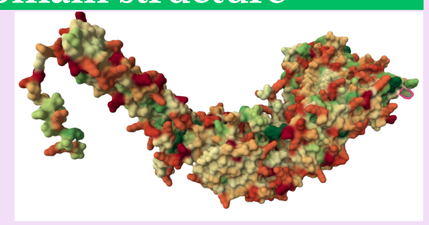


Figure 2B: 3D structure of Filamentous hemagglutinin N-terminal domain-containing protein predicted by Alphafold, based on the sequence from KBase. The protein is visualized as a surface-model and the colours represent hydrophobicity. Hydrophilic amino acids are red, hydrophobic amino acids are green. The amino acids sticking out of the surface are mostly hydrophilic. Of the amino acids in the tail (on the left), a large portion is hydrophobic (green/yellow).

In silico overexpression

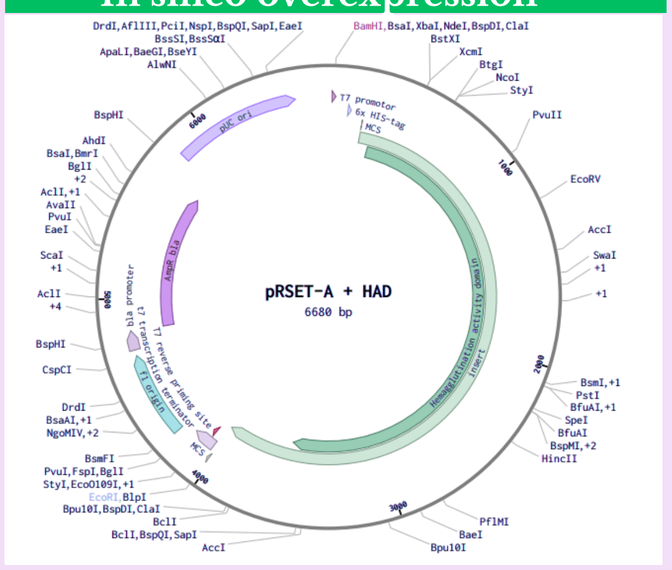


Figure 3: The pRSET-A with the hemagglutination activity domain inserted. It shows the gene within the MCS and downstream to the T7 promoter. The insert is longer than the actual gene as shown in the vector. Benchling was used to make the overexpression vector.

Conclusion

The goal of this case study was to use bioinformatics tools to study the organism Campylobacter jejuni and the hemagglutination activity domain protein. ORF26, which was discovered in DNA fragment 2 of set 8, codes for the hemagglutination activity domain (HAD) protein of campylobacter jejuni. This was confirmed to be the case by using multiple bioinformatics tools like NCBI-ORFfinder, NCBI-BLASTt and InterPro. Campylobacter jejuni is a pathogenic gram negative bacterium and is one of the most common causes of food poisoning. There was not much information online about the protein in this specific organism, so the protein was studied by comparing its function in multiple different organisms. HAD protein is found in many organisms and it has a function in adherence, this was discovered by protein analysis in interpro and the protein pathway in KEGG. Literature research showed that pathogens utilize this function to adhere to their host, which makes HAD protein a virulence factor. The Campylobacter jejuni can cause inflammation in the small intestine. It can also have a long term effect; for example the neurological disorder GBS. An in-silico design was made using benchling to overexpress the HAD gene in the inducible pRSET-A plasmid. This is important for therapeutic purposes like vaccines since the activity domain is an important potential target.

References

References poster

Logbook

Decoding *Staphylococcus aureus*: Properties, Mechanisms & Proteins

Authors: Jean Reichhold, Tim Holland, Merlin Steiner, Chinmay Jansen, Miguel Rodriguez, Renske de Vries

Abstract

This case study investigates an unknown DNA fragment that was traced back to the organism *Staphylococcus aureus (S. aureus)* using BLASTN. To further explore the genetic and functional contents of this fragment, various bioinformatics tools like NCBI BLAST, KBase, Benchling, etc. were used. The results are presented in this poster and documented in the corresponding logbook.

Introduction

Staphylococcus aureus is a facultative anaerobic, gram-positive bacterium that exists both as a commensal organism (microbiota, upper respiratory tract and skin) and as a pathogenic bacterium (bloodstream). Its circular genome ranges from 2,7 to 2,9 Mb and encodes for numerous virulence factors, including toxins. These toxins are particularly dangerous if the bacterium enters the bloodstream through a break in the skin, or in some cases through respiratory droplets.

Of particular concern are methicillin-resistant *S. aureus* (MRSA) strains, which are resistant to many available antibiotics. MRSA can cause severe hospital-acquired infections, such as aggressive skin infections and high fever, which are especially difficult to treat in immunocompromised individuals.

Overall, this makes *S. aureus*, especially MRSA strains, a significant pathogen in healthcare settings and a major challenge for clinical treatment.

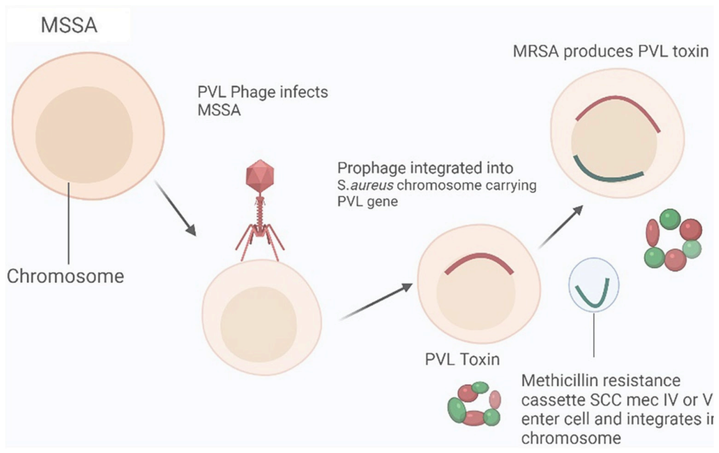


Fig. 1: *Depiction of an MSSA strain (Methicillin-sensitive S. aureus), that firstly get infected trough a PVL phage, carrying the methicillin resistance. During lysogenic infection, the PVL toxin integrates itself into the host chromosome, now called MRSA strain. This MRSA strain is then able to produce the PVL toxin, increasing its virulence and making it highly antibiotic-resistant.*

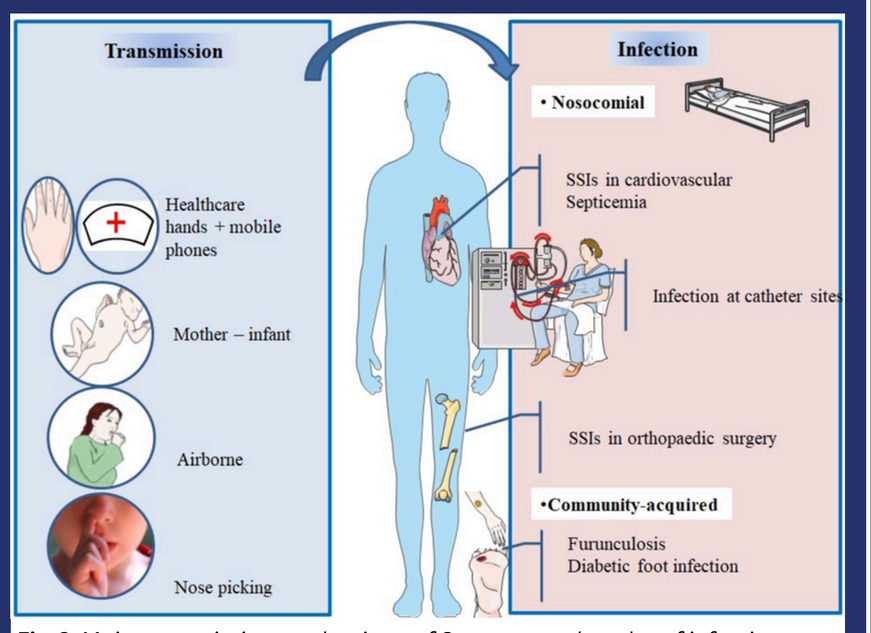


Fig. 2: *Main transmission mechanisms of S. aureus and modes of infections.*

Methods

1. **Sequence analysis:** *BLASTN* analysis identified the unknown DNA fragment as *S. aureus*. ORFs were predicted using *ORFfinder*, and *Prokka (KBase)* provided genomic annotation.
2. **Protein analysis:** DNA topoisomerase I was characterized using *Clustal Omega/Jalview* alignment across species and the *Protein DataBase/Google AlphaFold3* for structural analysis.
3. **Organism analysis:** Phylogenetic analysis was performed using *PhyloT/iTOL*. Operon structure and *KEGG pathway* proteins were determined with FACoP. Protein-protein interactions were visualized with the *STRING database*.
4. **In silico overexpression:** XerC gene specific primers with NcoI and HindIII restriction sites designed with *Primer-BLAST*, followed by in silico ligation of the XerC gene into the pBAD/His vector using *Benchling*.
5. **Pathogenicity assessment:** *Literature research* was performed, examining the pathogenicity, cause of infection, toxic mechanisms, transmission pathways, pathogen significance and global impact of *Staphylococcus aureus*.

Results

Description / Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident (%)	Acc. Len	Accession
Staphylococcus aureus subsp. aureus MSSA476	1425	1425	100%	0	99.85	689	Q6G9W3.1
Staphylococcus aureus subsp. aureus USA300	1423	1423	100%	0	99.56	689	Q2FHI8.2
Staphylococcus aureus	1422	1422	100%	0	99.56	689	Q06AK7.1
Staphylococcus aureus RF122	1419	1419	100%	0	99.42	689	Q2YXL8.2
Staphylococcus haemolyticus JCSC1435	1308	1308	100%	0	90.13	689	Q4L5V2.1
Staphylococcus epidermidis RP62A	1285	1285	100%	0	88.82	689	Q5HPU2.1
Staphylococcus saprophyticus subsp. saprophyticus	1219	1219	100%	0	86.5	688	Q49X35.1
Halalkalibacterium halodurans C-125	982	982	99%	0	67.05	690	Q9KA23.1
Bacillus subtilis subsp. subtilis str. 168	964	964	100%	0	67.44	691	P39814.1
Lactococcus lactis subsp. lactis Il1403	897	897	100%	0	61.73	710	Q9CG80.1

Fig. 3: *The NCBI table showing the PBLAST results based on the second largest ORF (53) , using the Swissprot database.*

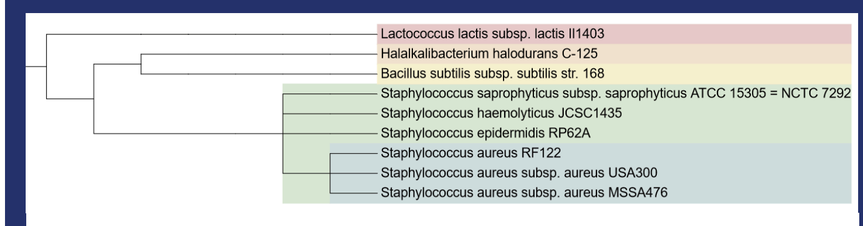


Fig. 4: *Phylogenetic tree of the top BLASTP hits specific to organisms encoding the Topoisomerase I enzyme.*

Type	Function	Aliases	Start	Strand	Length	Contig
CDS	Signal recognition particle receptor FtsY	ftsY	52	+	1,095	fragment_1
CDS	hypothetical protein		1,133	+	333	fragment_1
CDS	Succinate--CoA ligase [ADP-forming] subunit beta	sucC	10,898	+	1,167	fragment_1
CDS	Succinate--CoA ligase [ADP-forming] subunit alpha	sucD	12,086	+	909	fragment_1
CDS	hypothetical protein		13,318	+	873	fragment_1
CDS	DNA topoisomerase 1	topA	14,364	+	2,076	fragment_1
CDS	Methylenetetrahydrofolate--tRNA-(uracil-5-)-methyltransferase TrmFO	trmFO	16,595	+	1,308	fragment_1
CDS	Tyrosine recombinase XerC	xerC	18,319	+	897	fragment_1
CDS	ATP-dependent protease subunit HslV	hslV	19,212	+	546	fragment_1
CDS	Signal recognition particle protein	ffh	1,491	+	1,368	fragment_1

Fig. 5: *Top 10 results from KBase analysis of fragment 1, ordered by coding sequences (CDS) on the forward strand (+). The table provides the basis for the following investigated protein topoisomerase I and XerC gene.*

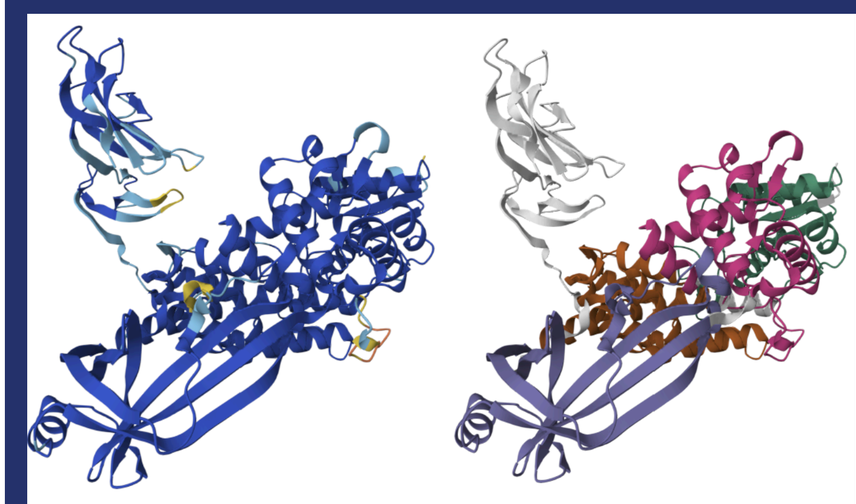


Fig. 6: *AlphaFold prediction of DNA topoisomerase I* ***(left)****. Model confidence scores: Dark blue (Very High pLDDT > 90), light blue (High pLDDT > 70), yellow (Low 70 > pLDDT > 50) and orange (Very low pLDDT < 50). The enzyme's 4 highly conserved central domains* ***(right)*** *differentiated by colors; green, pink, orange and purple.*

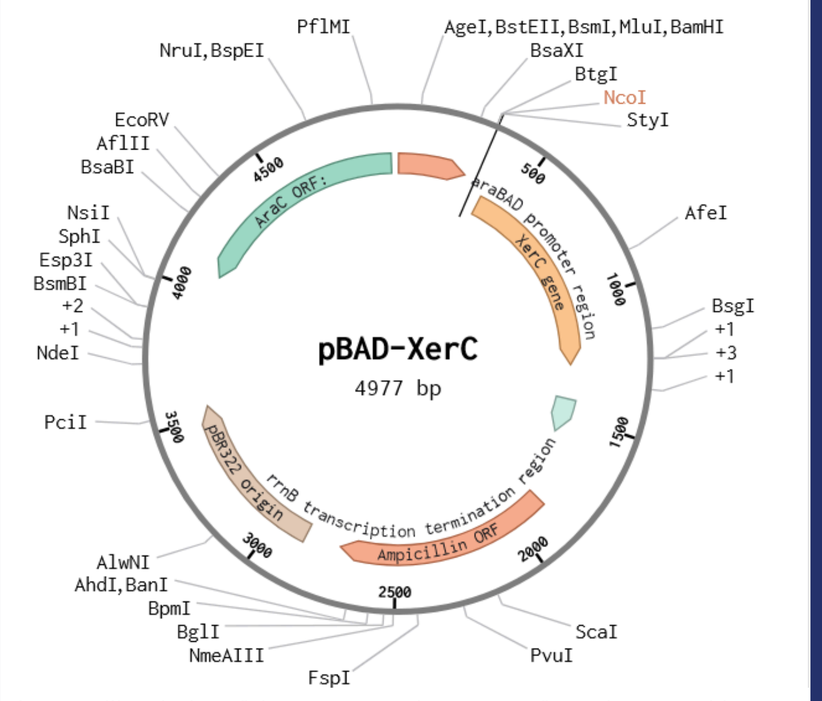


Fig. 7: *In silico design of the overexpression vector of tyrosine recombinase XerC. Using primers and restriction sites, the XerC gene was isolated from the genome, and the restriction sites were used to create complementary sticky ends in the plasmid / gene of interest for ligation.*

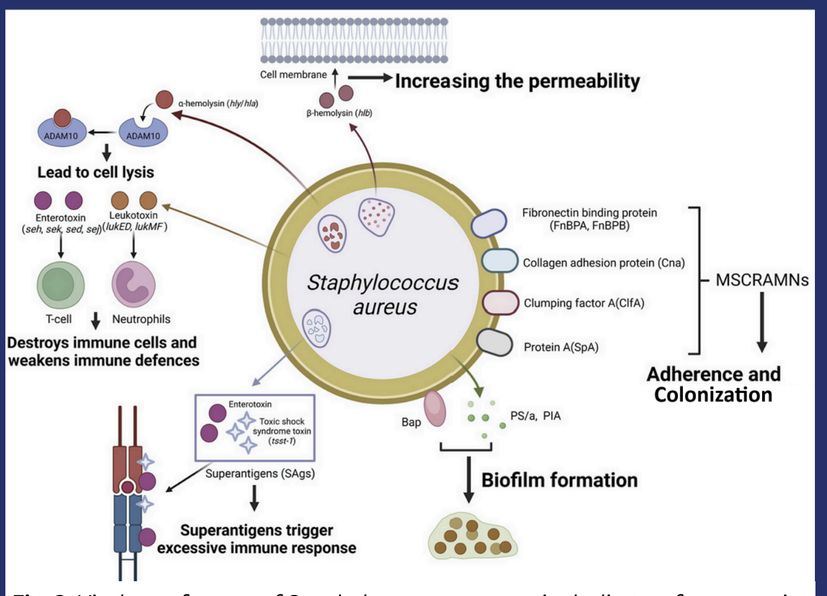


Fig. 8: *Virulence factors of Staphylococcus aureus, including surface proteins (MSCRAMMs), α-hemolysin–mediated cell lysis, leukotoxins and enterotoxins for immune evasion, superantigens triggering excessive immune responses, and secreted proteins promoting biofilm formation.*

***S. aureus* genome and pathogenic impact:**

- Its genome contains ~2830 genes, from which ~2750 are protein-coding.
- It is one of the leading causes of hospital acquired infections and deaths related to several toxins that are resistant to antibiotics.
- The global burden of MRSA has more than doubled over three decades: while MRSA infections resulted in 57,200 attributable deaths worldwide in 1990, this number increased to ~130,000 deaths in 2021.

Conclusion

The unknown sequence was traced back to Staphylococcus aureus. Using tools like NCBI blast, ORF finder, KBase, phyloT, iTOL, Benchling and Alphafold, related genes and proteins were analyzed.

Using literary research papers, we show the pathogenicity and virulence of S. aureus. Due to its surface proteins for adherence and colonization, and toxins like enterotoxin and leukotoxin, S. aureus can easily infect a host.

The specific protein retrieved from fragment 1 was DNA topoisomerase I, an enzyme that removes supercoils during replication.

The XerC gene was chosen for the in-silico design, to over express and generate primers for restriction analysis.

References

- Tong, X. et al. (2025) Virulence of bacteria causing mastitis in dairy cows: A literature review, MDPI. Available at: https://www.mdpi.com/2076-2607/13/1/167
- Siddiqui, A.H., K., Janak (2023) Methicillin-resistant Staphylococcus aureus, StatPearls [Internet]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK482221/
- Staphylococcus aureus basics (2024) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/staphylococcus-aureus/about/index.html
- Shoaib, M. et al. (2025) MRSA compendium of epidemiology, transmission, pathophysiology, treatment, and prevention within one health framework, Frontiers. Available at: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.1067284/full
- GBD 2021 Antimicrobial Resistance Collaborators (2024) Global burden of bacterial antimicrobial resistance 1990-2021: A systematic analysis with forecasts to 2050, Lancet (London, England). Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC11718157/
- Linz, M.S. et al. (2023) Clinical impact of staphylococcus aureus skin and soft tissue infections, Antibiotics (Basel, Switzerland). Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10044708/
- G. Ram, R.P Novick (2017) Staphylococcal Pathogenicity Islands-movers and shakers in the genomic firmament, Current opinion in microbiology. Available at: https://pubmed.ncbi.nlm.nih.gov/29100762/
- Zubair, S. et al. (2015) Complete genome sequence of Staphylococcus aureus, strain ilri_eymole1/1, isolated from a Kenyan dromedary camel - environmental microbiome, BioMed Central. Available at: https://environmentalmicrobiome.biomedcentral.com/articles/10.1186/s40793-015-0098-6
- Tong, S. Y. C., et al. (2015). Staphylococcus aureus infections: Epidemiology, pathophysiology, clinical manifestations, and management. Clinical Microbiology Reviews, 28(3), 603-661. https://doi.org/10.1128/CMR.00134-14
- Thingstad, T. F., et al. (2021). Pathogenicity and virulence of Staphylococcus aureus. Virulence, 12(1), 2175-2192. https://doi.org/10.1080/21505594.2021.1878688
- Tong, S. Y. C., et al. (2025). The virulence toolkit of Staphylococcus aureus: a comprehensive review of toxin diversity, molecular mechanisms and clinical implications. European Journal of Clinical Microbiology & Infectious Diseases, 44(5), 95-108. https://doi.org/10.1007/s10096-025-05148-y

Helicobacter Pylori

Pathology, proteins and characteristics

TEAM 10:
Lotte Kuivenhoven (S6167616)
Sterre Wiersma (S6005489)
Jasper Koenders (S5586364)
Sophie Rubingh (S6391753)
Alexander de Wit (S5401569)
Bent Stommels (S6111378)

rijksuniversiteit groningen

ABSTRACT

This case study centers on the detailed analysis of an unknown DNA fragment. The unknown DNA sequence was found to originate from the genome of ***Helicobacter Pylori***, using BLASTp. By using bioinformatic platforms such as NCBI ORFfinder, BLASTp, UniProt, and phylogenetic tree construction tools, we research the genetic properties of *H. pylori* and highlight the biological implications of a major **virulence factor** called **arginase**.

INTRODUCTION

Name:	**Helicobacter pylori**
Organism:	**Bacterium**
Category:	**Gram negative**
Shape:	**Helical and S-shaped**
Mobility:	**Flagella & chemotaxic action**
Incubation time:	**3 days**
Transmission:	**Oral - Oral, Fecal - Oral (trough water or food consumption)**

Table 1: General information of the organism *Helicobacter Pylori.*

In this case study, the unknown DNA sequence was found to originate from the genome of *Helicobacter Pylori*. *Helicobacter pylori* is a **spiral-shaped, gram-negative bacterium** that infects the lining of the **human stomach** and is one of the most common **chronic bacterial infections** worldwide, affecting more than half the global population. It is typically acquired during **childhood**, especially in areas with poor sanitation or crowded living conditions, and is often transmitted through person-to-person contact or contaminated food and water.
While many people with *H. pylori* infection have no symptoms, the bacterium can cause chronic gastritis, peptic ulcers, and, in rare cases, leads to stomach cancer or MALT lymphoma.

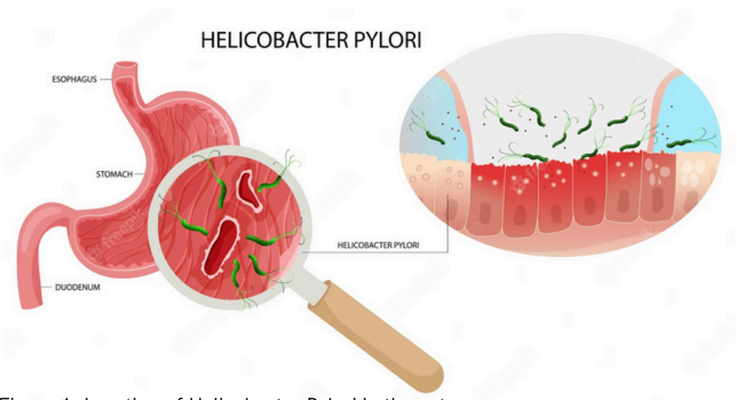


Figure 1: Location of Helicobacter Pylori in the gut

MATERIALS AND METHODS

Analysis of DNA fragments
NCBI ORF finder website is used to find the ORFs

- Settings used: standard settings of the site.

BLASTP is done to compare the protein sequence against a protein database to find similar sequences.

- Settings used: standard settings, the database selected: UniProtKB/Swiss-Prot(swissprot).

Literature research
To have a deeper understanding of *Helicobacter Pylori*, a literature study was done in respect to the virulence of the bacteria, the symptoms of infected individuals and the potential therapeutic effects Arginase might have.

Comparative protein study
The NCBI website was used to search for similar proteins. 20 similar protein sequences are selected. Selection requirements:

- Do not 100% Per. Ident.
- Do not have same AA length

Clustral Omega webserver is used to form a Guide tree and phylogenetic tree, to later use in iTOL.The tool called "iTOL (interactive tree of life)" is used to build a phylogenetic tree. The colour coding is done based on taxonomy (the three domains), this information was found on NCBI.

Detailed protein analysis
To start, the sequence of arginase was put in InterPro to identify domains, biological processes and functions. Furthermore, the STRING database was to visualize the protein network of arginase. At last to further visualize the protein in 2D and 3D PDB and Alphafold were used.

RESULTS

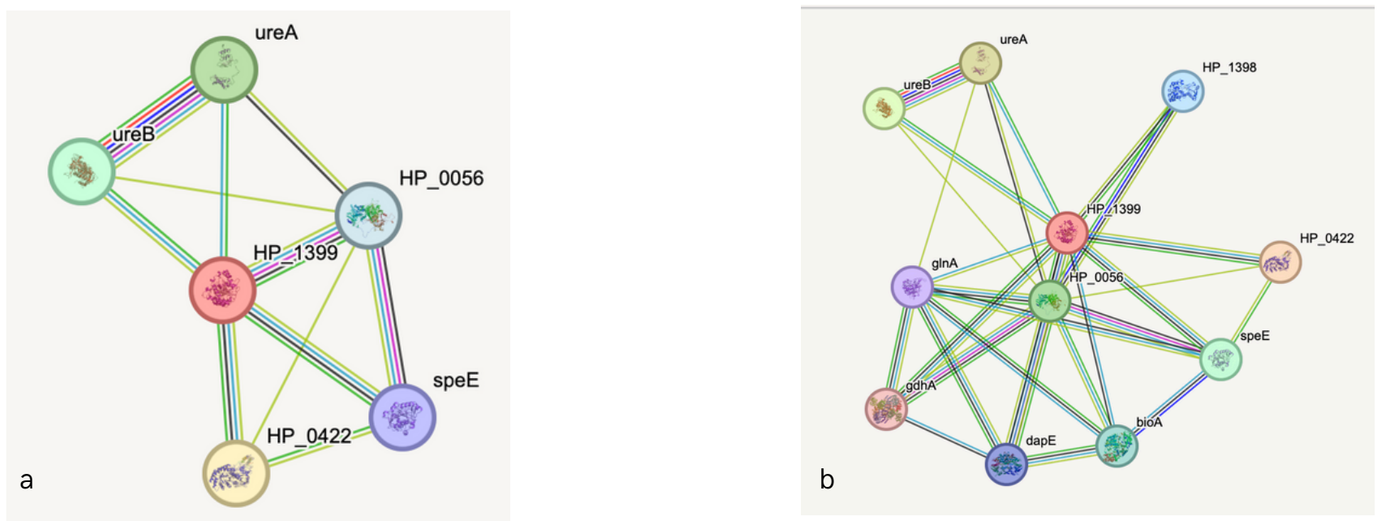


Figure 2: Result of STRING network of the protein Arginase. Interaction network of Arginase (b) and the smaller interactions (a). The different coloured strings show different interactions.

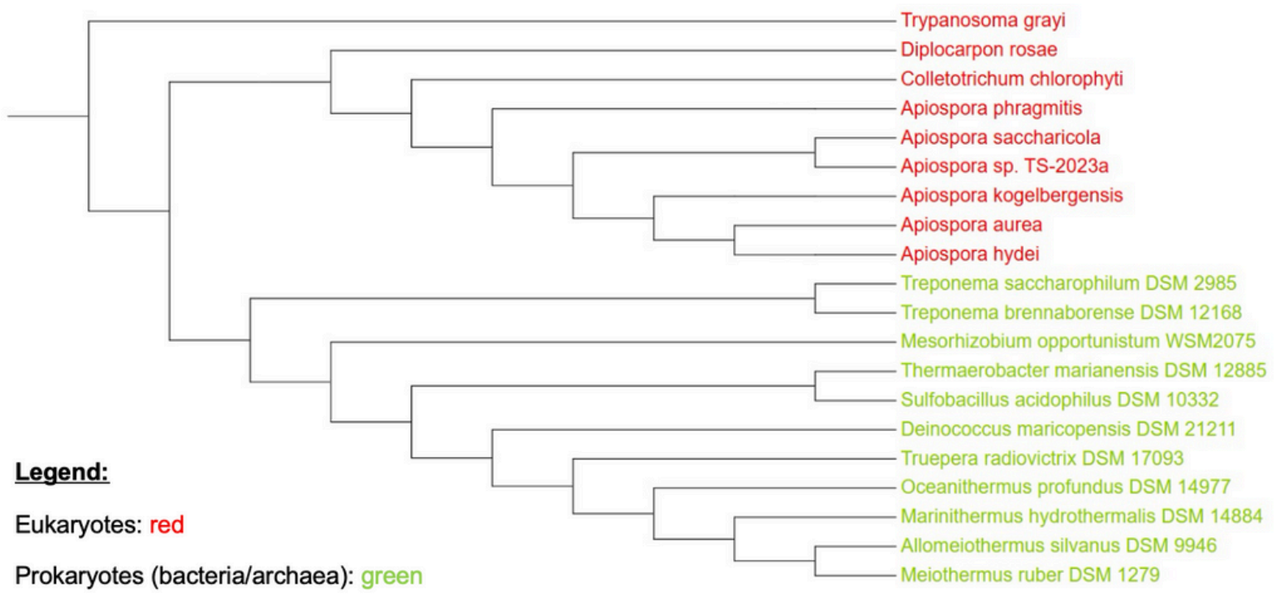


Figure 3: Pylogentic tree results of the organisms who share the protein Arginase. Coloured by their taxonomy

From Interpro, the following information on the protein retrieved:

- **Domains:**
  - Arginase/deacetylase domain (superfamily SSF52768)
  - Ureohydrolase domain
- **Biological processes:**
  - arginine metabolic process → The chemical reactions and pathways involving arginine, 2-amino-5-(carbamimidamido)pentanoic acid.
  - l-arginine + h2o > urea + l-ornithine
- **Molecular function:**
  - arginase activity
  - metal ion binding → manganese ion binding
- **Families:**
  - Gene locus:-Located at the rocF gene
  - rocF-arginase → ureohydrolase superfamily

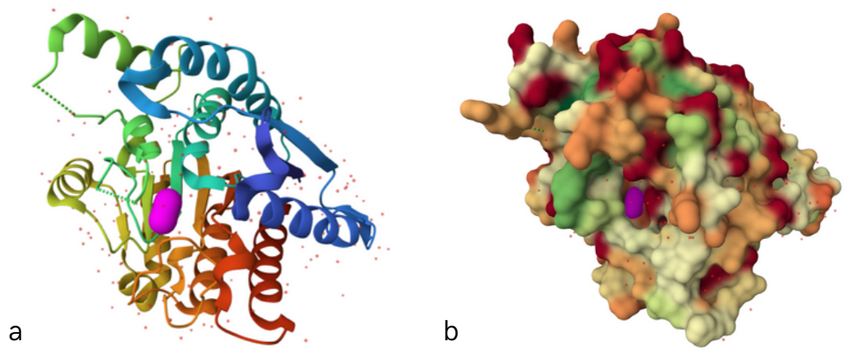


Figure 4: 3D structure (a) and molecular surface (b) representation generated with PDB of the protein Arginase.

CONCLUSIONS

The analysis shows that arginase is a **highly conserved** protein found in both prokaryotes and eukaryotes, indicating its essential functional role (figure 3). In *Helicobacter pylori*, arginase belongs to the **rocF family** and is part of the **ureohydrolase superfamily**. It converts L-arginine into urea and L-ornithine, reducing host nitric oxide production and **aiding immune evasion**. The enzyme is manganese-dependent and has a well-defined catalytic pocket formed by alternating α-helices and β-sheets. Encoded by the RocF gene at the **HP_1399 locus**, it is active in the cytoplasm. STRING analysis further suggests that arginase is part of a broader metabolic network crucial for ***H. pylori*** survival (figure 2).

REFERENCES

1. Öztekin, M., Yılmaz, B., Ağagündüz, D., & Capasso, R. (2021). Overview of Helicobacter pylori Infection: Clinical Features, Treatment, and Nutritional Aspects. *Diseases*, *9*(4), 66. https://doi.org/10.3390/diseases9040066
2. *H. Pylori infection*. (2025, 25 september). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/21463-h-pylori-infection
3. Public Health Agency of Canada. (2024, 25 november). *Helicobacter pylori: Infectious substances pathogen safety data sheet*. Canada.ca. https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/helicobacter-pylori.html
4. Zhang, J., Zhang, X., Wu, C., Lu, D., Guo, G., Mao, X., Zhang, Y., Wang, D., Li, D., & Zou, Q. (2011). Expression, Purification and Characterization of Arginase from Helicobacter pylori in Its Apo Form. *PLoS ONE*, *6*(10), e26205. https://doi.org/10.1371/journal.pone.0026205
5. Gobert, A. P., McGee, D. J., Akhtar, M., Mendz, G. L., Newton, J. C., Cheng, Y., Mobley, H. L. T., & Wilson, K. T. (2001). Helicobacter pyloriarginase inhibits nitric oxide production by eukaryotic cells: A strategy for bacterial survival. *Proceedings Of The National Academy Of Sciences*, *98*(24), 13844-13849. https://doi.org/10.1073/pnas.241443798
6. Azizian, H., Bahrami, H., Pasalar, P., & Amanlou, M. (2010). Molecular modeling of Helicobacter pylori arginase and the inhibitor coordination interactions. *Journal Of Molecular Graphics And Modelling*, *28*(7), 626-635. https://www.sciencedirect.com/science/article/abs/pii/S1093326309001855

Living positive:

Comparative Bioinformatics of the NhaA Antiporter in *H. pylori*

E. Bosman[1] (S5605725); **E. Koenjer**[1] (S5688221); **J. Marnink**[1] (S6520480); **S. Oloeru**[1] (S5175712); **J. Tatuhey**[1] (S6238661); **A. Verdonck**[1] (S5596874).

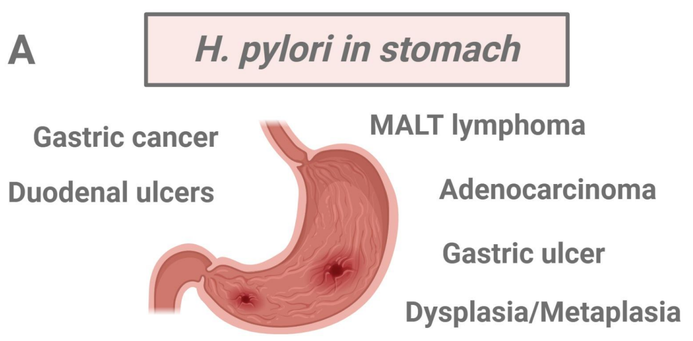
[1]Rijksuniversiteit Groningen, Bioinformatics course, WBBY002-05, Groningen, The Netherlands

ABSTRACT

This case study analyzes an unknown DNA fragment from *Helicobacter pylori* to explore its genetic composition and biological relevance. Sequence analyses identified conserved genes involved in acid adaptation, including the Na^+/H^+ antiporter (NhaA), which supports pH regulation and survival in acidic environments. Protein and phylogenetic analyses revealed structural differences from *Escherichia coli*, suggesting evolutionary adaptation to gastric conditions. An in silico pUC19 plasmid incorporating the NhaA gene was designed to model expression and functional studies. These results provide a framework for future experimental validation and give an insight for potential applications.

INTRODUCTION

Helicobacter pylori is a spiral-shaped, gram-negative bacterium that colonizes the human stomach and infects nearly 44% of the global adult population (Chen et al., 2024). *H. pylori* infection is a major public health concern, strongly associated with gastritis, peptic ulcer disease, and gastric cancer (Kao et al. 2016). Transmission commonly occurs during childhood through oral–oral or fecal–oral routes, particularly in regions with poor sanitation and overcrowding (Li et al., 2023). In this case study, we examine a *Helicobacter pylori* DNA fragment to identify its encoded proteins and explore the characteristics and evolutionary relationships of the Na^+/H^+ antiporter.



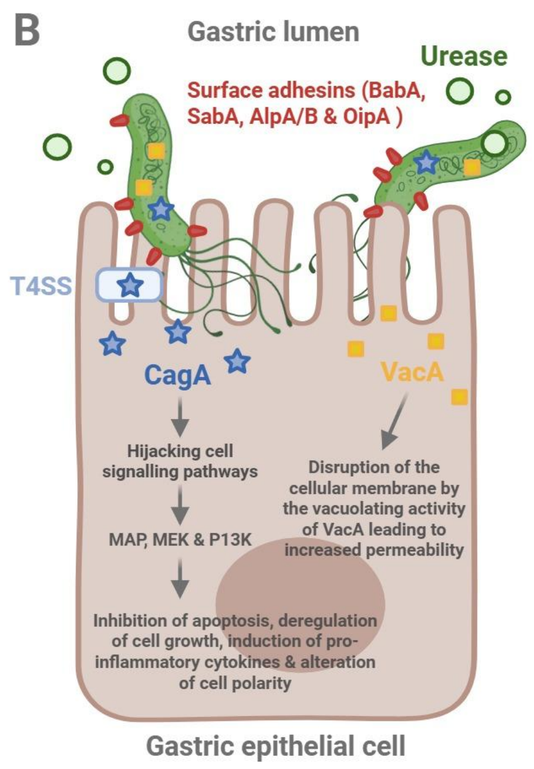


Figure 1 | *Helicobacter pylori (H. pylori)* pathogenesis and associated gastric diseases: A. Diagram showing *Helicobacter pylori*-associated gastric disorders, including duodenal and gastric ulcers, adenocarcinoma, MALT lymphoma, and dysplasia/metaplasia arising from chronic infection (Ali & Alhussani, 2024). **B.** Overview of *Helicobacter pylori* interaction with gastric epithelial cells, highlighting key virulence factors: CagA, VacA, and surface adhesins, affecting host signaling, membrane integrity, and inflammation (Created with Biorender).

METHODS

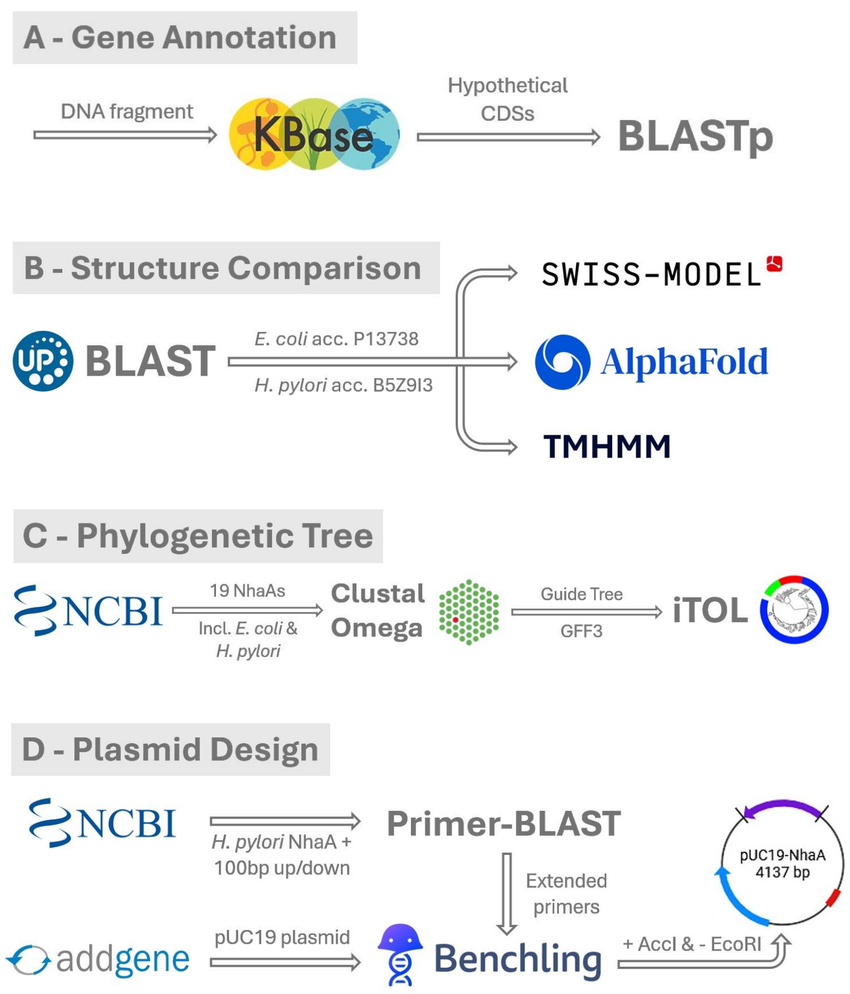


Figure 2 | Overview of the bioinformatics case study workflow: A. Gene annotation was performed with KBase and BLASTp to identify coding sequences and their function. **B.** Protein structures were predicted and compared through BLAST, SWISS-MODEL, AlphaFold, and TMHMM. **C.** Phylogenetic relationships among NhaA homologs were inferred with Clustal Omega and visualized in iTOL. **D.** Finally, *H. pylori* NhaA was cloned into the pUC19 vector using AccI and EcoRI for plasmid construction.

RESULTS

Table 1 | Genomic features identified within the *Helicobacter pylori* DNA fragment using Kbase annotation tools.

Function	Start	Strand	Length
Transcription-repair-coupling factor	3,540	-	3000
Bactofilin family protein	3,858	-	315
Peptidoglycan DD-metalloendopeptidase Csd1	4,781	-	909
M23 family metallopeptidase	5,743	-	927
Dihydrofolate synthase/folylpolyglutamate synthase	6,918	-	1176
LPS assembly lipoprotein LptE	7,420	-	513
Leucine--tRNA ligase	9,837	-	2421
Protein translocase subunit SecF	11,166	-	972
Protein translocase subunit SecD	12,752	-	1578
Sec translocon accessory complex subunit YajC	13,045	-	294
Na(+)/H(+) antiporter NhaA	14,409	-	1317
RecBCD enzyme subunit RecB	17,228	-	2796
30S ribosomal protein S2	17,455	-	843
Elongation factor Ts	18,297	+	1068

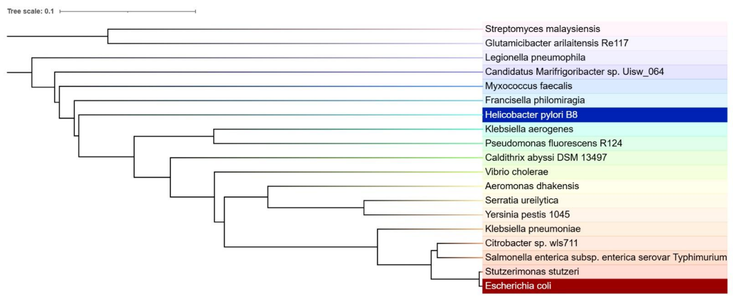


Figure 3 | Phylogenetic tree of NhaA proteins in different species. H. pylori, E. coli and 17 other random NCBI search hits were used to create the tree. The branch lengths represent evolutionary distance, indicated by a scale on top of the tree. Different colours separate the different species. The proteins of interest, from *H. pylori* and *E. coli*, are given a bolder colour.

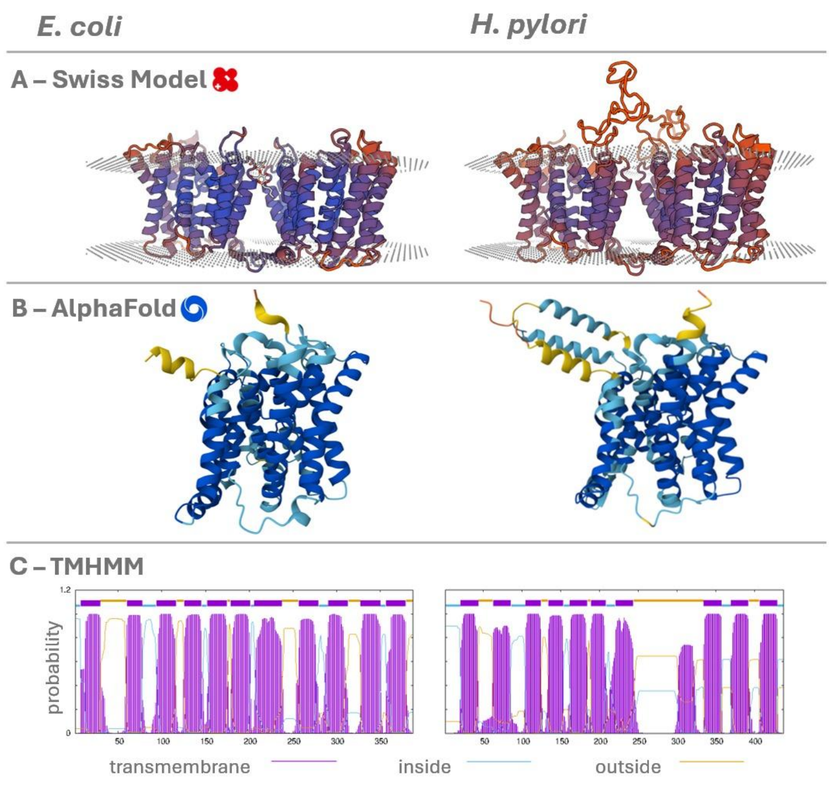


Figure 4 | Structural comparison of NhaA in *E. coli* and *H. pylori*. A. Swiss-Model predicted 3D structures of dimerized NhaA with membrane visualization (grey) **B.** AlphaFold-predicted NhaA structures colored by model confidence (pLDDT): dark blue (very high) to red (very low). **C.** TMHMM topology (fuchsia: transmembrane; blue: inside; orange: outside).

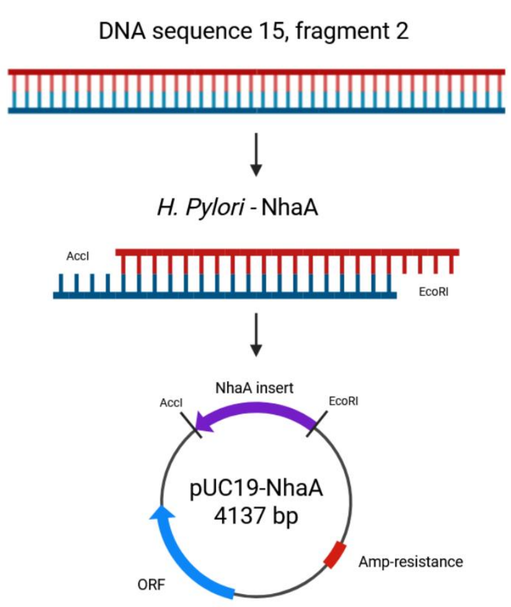


Figure 5 | pUC19 plasmid with *Helicobacter Pylori* pH-dependent sodium/proton antiporter insert. Circular map of the 4137 bp recombinant pUC19 plasmid showing key elements, including the origin of replication (ori), ampicillin resistance gene (ampR), and multiple cloning site (MCS). The pH-dependent sodium/proton antiporter gene from *Helicobacter pylori* was inserted at the lac-operon site using the **AccI** and **EcoRI** restriction enzymes (Created with BioRender).

CONCLUSIONS

The genomic fragment was identified as *Helicobacter pylori* through BLAST and gene annotation, revealing genes such as the Na^+/H^+ antiporter (HpNhaA), Sec translocon components, and DNA repair enzymes. Structural comparison showed that HpNhaA contains an inserted extracellular domain (aa 254–294) absent in *E. coli*, likely linked to its function across a broader pH range (6.5–8.5) in the acidic gastric niche (Calinescu et al. 2014). Phylogenetic analysis confirmed its evolutionary divergence from neutral-pH bacteria, reflecting long-term host adaptation (Atherton et al. 2009). A plasmid design was proposed to express *H. pylori* NhaA in *E. coli* for functional validation of these acid-tolerance features (Marcus et al. 2018).

Potential applications:

- Probiotics – so they may pass the gastric digestion
- Agriculture – so they may adapt to the lowering pH levels in our agricultural soils
- Biotech – so they may survive harsher conditions in bioreactors

REFERENCES

Ali, A., & AlHussaini, K. I. (2024). *Helicobacter pylori: A contemporary perspective on pathogenesis, diagnosis and treatment strategies*. *Microorganisms, 12*(1), 222. https://doi.org/10.3390/microorganisms12010222

Atherton, J. C., & Blaser, M. J. (2009). Coadaptation of *Helicobacter pylori* and humans: Ancient history, modern implications. *Journal of Clinical Investigation, 119*(9), 2475–2487. https://doi.org/10.1172/JCI38605

Călinescu, O., Danner, E., Böhm, M., Hunte, C., & Fendler, K. (2014). Species differences in bacterial NhaA Na^+/H^+ exchangers. *FEBS Letters, 588*(14), 2511–2516. https://doi.org/10.1016/j.febslet.2014.05.066

Chen, Y.-C., Malfertheiner, P., Yu, H.-T., Kuo, C.-L., Chang, Y.-Y., Meng, F.-T., Wu, Y.-X., Hsiao, J.-L., Chen, M.-J., Lin, K.-P., Wu, C.-Y., Lin, J.-T., O'Morain, C., Mégraud, F., Lee, W.-C., El-Omar, E. M., Wu, M.-S., & Liou, J.-M. (2024). Global prevalence of *Helicobacter pylori* infection and incidence of gastric cancer between 1980 and 2022. *Gastroenterology, 166*(4), 1–15. https://doi.org/10.1053/j.gastro.2023.12.022

Kao, C.-Y., Sheu, B.-S., & Wu, J.-J. (2016). *Helicobacter pylori* infection: An overview of bacterial virulence factors and pathogenesis. *Biomedical Journal, 39*(1), 14–23. https://doi.org/10.1016/j.bj.2015.06.002

Li, Y., Choi, H., Leung, K., Jiang, F., Graham, D. Y., & Leung, W. K. (2023). Global prevalence of *Helicobacter pylori* infection between 1980 and 2022: A systematic review and meta-analysis. *The Lancet Gastroenterology & Hepatology, 8*(6), 555–567. https://doi.org/10.1016/S2468-1253(23)00070-5

Marcus, E. A., Sachs, G., & Scott, D. R. (2018). Acid-regulated gene expression of *Helicobacter pylori*: Insight into acid protection and gastric colonization. *Helicobacter, 23*(3), e12490. https://doi.org/10.1111/hel.12490

Yersinia Pestis: Pathogenesis and evolution

S-numbers: S5634687, S5775361, S5949297, S6063365 and S5826489

Bioinformatics

Abstract

Characterization of an unknown DNA fragment is important for understanding its pathogenesis and evolution. In this case study, a DNA fragment was analyzed and the organism of origin was determined to be ***Yersinia pestis***. It is the causative agent of plague, which evolved from *Yersinia pseudotuberculosis* ~5,000 years ago gaining increased virulence and transmission.

Introduction

In this study *Y.pestis* was identified from an unknown DNA fragment. Then, the Invasin pseudogene on this fragment was chosen to be studied further using literature research, gene annotation, multiple sequence alignment, a phylogenetic tree and analysis of the 3D structures.

The gram-negative nonmotile *Y.pestis* caused a total of 3 deadly plagues, including the black death killing around 1/3 of Europe (Barbieri, et all., 2020). The blood-borne disease normally occurs in rodents but could spread to humans by flea bites. According to experiments, *Y.pestis* descended from *Yersinia pseudotuberculosis, an enteropathogen,* around 5,000 years ago. Horizontal DNA transfer likely increased *Y. pestis'* virulence, therefore certain genes were no longer needed, like the Invasin gene. Those were lost (Parkhill, et all., 2001). Two adhesins are now expressed at increased levels; PsaA and the Ail protein, an outer membrane protein which mediates host cell attachment, provides resistance to the complement system of the host and plays an essential role in the administration of virulence factors into the host cells (Seabaugh & Anderson, 2024).

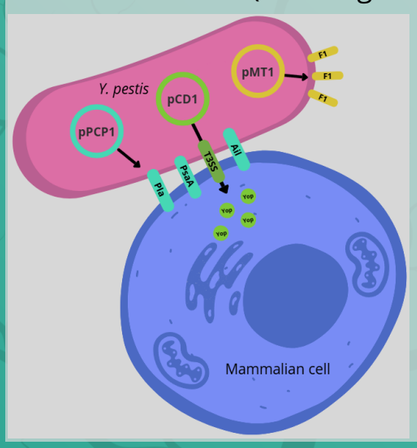


Figure 1. Pathogenesis of *Y. pestis*. Ail, PsaA and Pla are adhesins that keep the cell attached to the host and help evade the immune system. F1 forms small pili that aid in adherence and immune evasion. Through T3SS, Yops are injected that act on many pathways and establish an infection (Seabaugh & Anderson, 2024).

Materials and methods

BLASTN and ORFfinder:
Identification of *Y. pestis* via UniprotKB/Swiss-Prot and ClusteredNR database. To predict ORFs and identify genes via ORFfinder, these settings were used: Genetic code 11, start codon "ATG" and alternative initiation codons.

KBase and BLASTX:
Gene annotation by KBase database (Prokka) to identify genes and to confirm the Invasin gene and BLASTX against both the UniprotKB/Swiss-Prot and ClusteredNR database.

Multiple Sequence Alignment (MSA) and phylogenetics:
10 Yersinia species, including *Y. pestis*, selected from the ClusteredNR database and a Clustal Omega MSA was performed. Phylogenetic tree generated in iTOL (Mode = rectangular, Branch lengths = ignore)

3D protein structure analysis :
RCSB PSB was used to determine 3D structures of Invasin and Ail.

Results

Function	Start (nt)	Strand	Length (nt)
Protein YebF	325	+	411
Protease 2	1064	+	2052
Proline iminopeptidase	3837	+	951
DNA pol. III subunit theta	5077	-	231
Bacterial non-heme ferritin	5478	+	510
Protein YobA	6393	+	387
IMP YebZ	6782	+	885
Flagellar biosynthetic protein FlhB	9937	+	1161
Flagellar biosynthetic protein FlhA	11097	+	2079
Flagellar protein FlhE	13175	+	423
Invasin	14255	+	1011
Transposase IS200G	15255	+	459
Invasin	15803	+	2073
Flagella synthesis protein FlgN	18502	-	438
Negative regulator of flagellin synthesis	18826	-	303

Table 1. Results of Kbase. Annotation of the genes on the DNA fragment of *Y. pestis* using Kbase database. Hypothetical proteins were left out. Multiple Flagellar genes were located and two annotations of the Invasin gene.



Description	Scientific name	Percentage identity	E-value	Accession number
Invasin	Y. pseudotuberculosis	98.83%	0.0	P11922.2
Invasin	Y. enterocolitica	46.26%	1e-123	P19196.1
Intimin	Escherichia coli	32.10%	3e-39	P43261.2
Intimin	E. coli	32.51%	6e-39	O31000.1
Intimin	E. coli	31.60%	3e-38	P19809.2
Intimin	Citrobacter freundii	28.81%	1e-33	Q07591.1
Inverse autotransporter adhesin	E. coli	36.89%	2e-32	P76347.4

Table 2. Results of BLASTX. 6 best hits for ORF85, located via ORFfinder. Sorted by E-value.

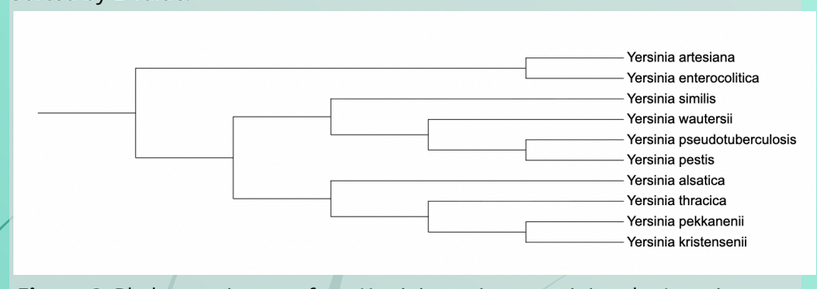


Figure 2. Phylogenetic tree of ten *Yersinia* strains containing the Invasin gene. Constructed with iToL. *Y. pestis* is most closely related to *Y. pseudotuberculosis*.

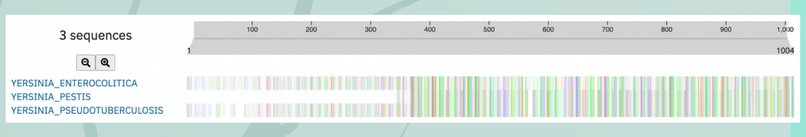


Figure 3. Clustal Omega Multiple Sequence Alignment of *Y. pestis* and two related strains with a working Invasin gene (1-1004 AA)

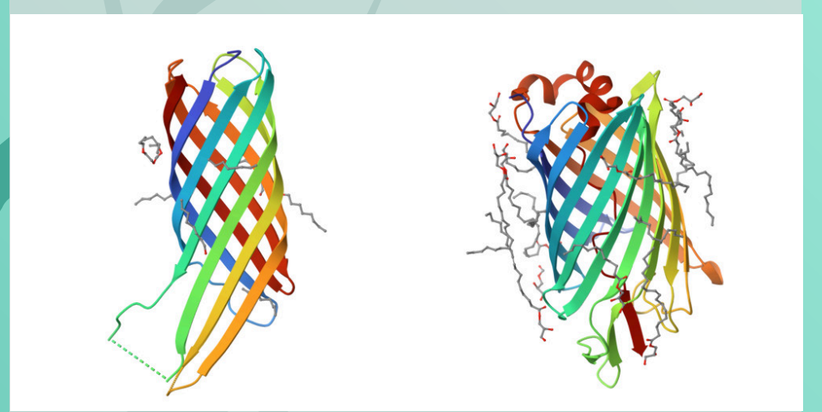


Figure 4. X-ray crystal structure of Ail protein β-barrel found in *Y. pestis* (left) and the X-ray crystal structure of Invasin from *Y. pseudotuberculosis* (right) derived from RCSB PDB

Conclusion

In this case study, we identified the Invasin pseudogene on the *Y. pestis* DNA fragment and revealed its similarity to Invasin in *Y. pseudotuberculosis* and other related species. Our gene annotation and phylogenetic analysis results showed that the Invasin gene in *Y. pestis* is disrupted due to an IS200G insertion, supporting the literature. The highly expressed Ail protein used in *Y. pestis*, which serves a similar function to Invasin, showed differences in 3D protein structure, explaining their different roles in adhesion and invasion. Together, these findings illustrate the virulence strategies in *Y. pestis*, specifically looking at the Invasin and Ail protein.

References

- Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., & Carniel, E. (1999). Yersinia pestis , the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proceedings of the National Academy of Sciences, 96(24), 14043–14048.
- Bank, R. P. D. (n.d.). RCSB PDB - 3QRA: The crystal structure of Ail, the attachment invasion locus protein of Yersinia pestis.
- Bank, R. P. D. (n.d.). RCSB PDB - 4E1T: X-ray crystal structure of the transmembrane beta-domain from invasin from Yersinia pseudotuberculosis.
- Barbieri, R., Signoli, M., Chevé, D., Costedoat, C., Tzortzis, S., Aboudharam, G., Raoult, D., & Drancourt, M. (2020). Yersinia pestis: the Natural History of Plague. Clinical Microbiology Reviews, 34(1).
- Deng, W., Burland, V., Plunkett, G., Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plano, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., . . . Perry, R. D. (2002). Genome Sequence of Yersinia pestis KIM. Journal of Bacteriology, 184(16), 4601–4611.
- Felek, S., & Krukonis, E. S. (2008). TheYersinia pestisAil protein mediates binding and YOP delivery to host cells required for plague virulence. Infection and Immunity, 77(2), 825–836.
- Hinnebusch, B. J., & Erickson, D. L. (2008). Yersinia pestis Biofilm in the Flea Vector and Its Role in the Transmission of Plague. Current Topics in Microbiology and Immunology, 229–248.
- Isberg, R. R., Hamburger, Z., & Dersch, P. (2000). Signaling and invasin-promoted uptake via integrin receptors. Microbes and Infection, 2(7), 793–801.
- Ke, Y., Chen, Z., & Yang, R. (2013). Yersinia pestis: mechanisms of entry into and resistance to the host cell. Frontiers in Cellular and Infection Microbiology, 3.
- Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., Baker, S., Basham, D., Bentley, S. D., Brooks, K., Cerdeño-Tárraga, A. M., Chillingworth, T., Cronin, A., Davies, R. M., Davis, P., Barrell, B. G. (2001). Genome sequence of Yersinia pestis, the causative agent of plague. Nature, 413(6855),
- Seabaugh, J. A., & Anderson, D. M. (2024). Pathogenicity and virulence of Yersinia. Virulence, 15(1).
- Simonet, M., Riot, B., Fortineau, N., & Berche, P. (1996). Invasin production by Yersinia pestis is abolished by insertion of an IS200-like element within the inv gene. Infection And Immunity, 64(1), 375–379.
- Tsang, T. M., Wiese, J. S., Felek, S., Kronshage, M., & Krukonis, E. S. (2013). Ail Proteins of Yersinia pestis and Y. pseudotuberculosis Have Different Cell Binding and Invasion Activities. PLoS ONE, 8(12), e83621.

From Genome to Host
Bioinformatics Insights into Enterococcus faecalis

university of groningen
faculty of science and engineering

Raggers J. (S5861837)
Sivaram Murugan N. (S6012604)
Avram M. G. (S5977290)
Lupsa S. (S5985293)
Felföldi K. A. (S5587069)
Mic A. (S5251060)

The analysis included identifying Enterococcus faecalis from a DNA fragment, annotating genes and proteins, analyzing metabolic pathways, protein structures, overexpression vectors, domains and interactions, studying pathogenic mechanisms, human immune responses, phylogeny, gene regulation and genomic islands, transmission, and proposing treatment strategies.

Introduction

Organism information

E. faecalis is a gram positive, coccus shaped lactic acid bacteria, which thrives in the nosocomial environment and is linked to various infections such as UTIs, endocarditis and bacteraemia. [1]

E faecalis displays multiple traits that increase its virulence.

- **Adherence to the host tissue** - 4 specific types
- **Capsular polysaccharides** - evade phagocytosis
- **Sex pheromones** - conjugative plasmid system
- **Etracellular superoxide** - disruptive oxygen radical
- **Hydrolytic enzymes** - Gelatinase and Hyaluronidase
- **Cytolysis** - forms pores in the cytoplasmic membrane of target cells
- **Biofilm formation** - aids survival in harsh conditions



Figure 1: Experimental structure of Glutamine-fructose-6-phosphate transaminase protein from PBD

Glutamine-fructose-6-phosphate transaminase (isomerizing)

The protein Glutamine-fructose-6-phosphate transaminase (isomerizing) Figure 1, is encoded by ORF14 and takes part in carbohydrate metabolism as an enzyme.

Role in virulence

E. faecalis can mitigate the host immune response by downregulating its carbohydrate metabolism, which also helps in facilitating its own survival in macrophages. [2] As Glutamine-fructose-6-phosphate transaminase (isomerizing) is a carbohydrate derivative and has a role in biosynthetic processes it is also affected by this downregulation. [3]

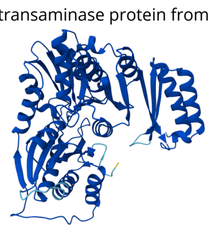
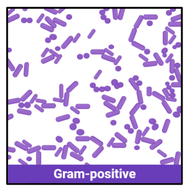


Figure 2: Predicted structure of Phosphoglucosamine mutase from AlphaFold



Pathogenicity islands

Pathogenic islands: IslandViewer 4 analysis of E. faecalis ATCC 29212, against reference genome of E. faecalis V583, detected genomic islands elsewhere, but none within the 84–104 kb fragment, confirming no virulence or resistance genes. [4]

Genes: Contains 15 annotated “housekeeping” and metabolic genes linked to cell-wall biosynthesis, amino-sugar metabolism, energy conversion, stress regulation, nutrient transport, and signaling control. [5]

GC Content: The 20 kb fragment shows a 38.1% GC content, matching E. faecalis’s genome average (~37.5%), indicating native chromosomal origin without horizontal gene transfer evidence.

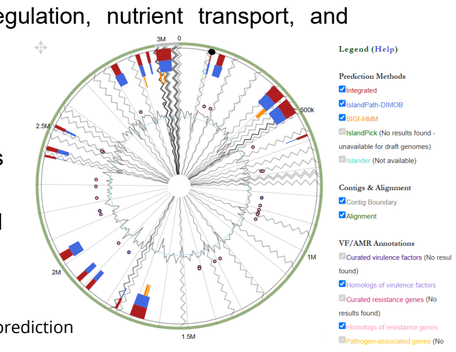
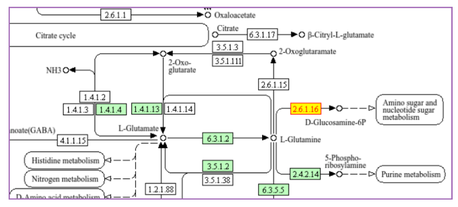


Figure 3: GI prediction

Metabolic Pathways

KEGG orthology: mapping links six genes to metabolic and regulatory pathways, showing evolutionary conservation. Most enzymes participate in shared bacterial processes like carbohydrate, nitrogen, and tRNA metabolism, while cdaA’s c-di-AMP signaling is uniquely restricted to Gram-positive bacteria. [3]

Figure 4: The enzyme catalyzes the first step in hexosamine metabolism, converting fructose-6-phosphate into glucosamine-6-phosphate using glutamine as a nitrogen source leading to Amino sugar and Nucleotide sugar biosynthesis.

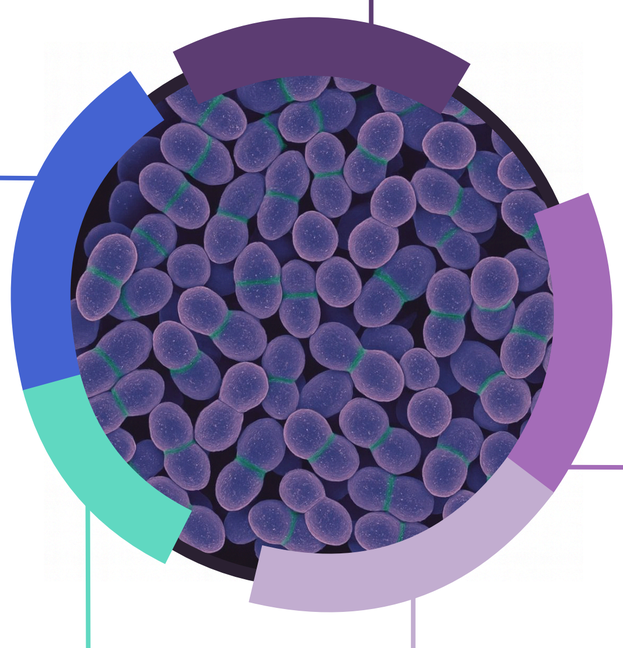


Enzymes: glmS and glmM for peptidoglycan biosynthesis, glnA and nifJ for nitrogen and energy metabolism, cdaA for signaling, and miaA and hflX for translation stability—indicating conserved core metabolic functions. [3]

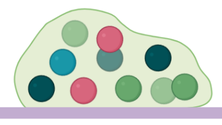
Methods

GenBank annotation – Verified gene functions.
KEGG mapping – Linked genes to pathways.
IslandViewer 4 – Detected genomic islands.
GC content – Calculated nucleotide composition.
Orthology analysis – Compared evolutionary conservation.
Enzyme classification – Assigned EC/KO roles.
Comparative genomics – Aligned with reference genome.
Benchling - Overexpression vector and related analysis
ORFfinder - Can find ORFs in a DNA sequence
UniProt - Information on proteins
InterPro - Information of protein domains and functions
BLAST (NCBI) – Identified homology and annotations of DNA and protein sequences.
Literature study – Reviewed E. faecalis research.









Overexpression vector

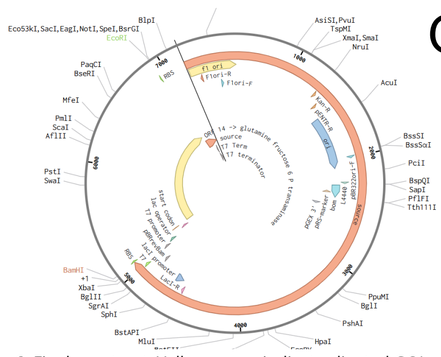


Figure 6: Final construct. Yellow arrow indicates ligated GOI

Relevance

Glutamine 6P transaminase is involved in carbohydrate metabolism. The overexpression leads to a decreased virulence, as the pathogens ability to downregulate its metabolism and avoid host detection is hindered. [2]

Experimental design

The glutamine *fructose 6P transaminase*, encoded by ORF14 in the target sequence was amplified in silico via PCR and cloned into and overexpression vector.

The GOI extends in the original sequence from bp 3396 to bp 1588, on the negative strand. Two primers were designed to flank the GOI on either side:
Forward primer: EcoRI restriction site
Backward primer: BamHI restriction site

In silico PCR was performed to obtain the cloned GOI. The pET vector was chosen and the final construct (figure 6) was successfully assembled through a digestion and ligation protocol, using BamHI and EcoRI as restriction sites.

Transmission and infection

Cause of infection

The overgrowth of the bacterium in the gut increases the probability of it translocating to the bloodstream and to the other body parts. [15]

Transmission

The transmission of the bacterium can happen through both direct contact and indirect contact. [16]

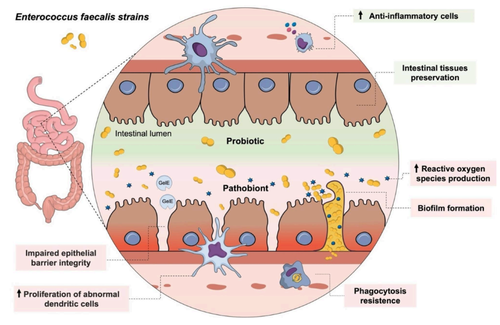
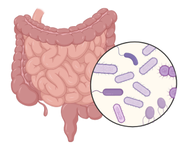


Figure 5: Dual characterecteristics of E. faecalis depending on enviroment

Mechanisms of Translocation

- **Engulfing** of the pathogen by macrophages:
  - Once engulfed, the bacterium is transported across the wall of the intestine & into the underlying lymph system. Failure to kill the bacterium can lead to a systemic spread.
- **Diffusion** of the Enterococci from the gut into the bloodstream:
  - In a healthy individual, the innate system easily destroys the pathogen, but in an individual with a weakened immune system, the bacterium may not be killed which can lead to contamination of other body parts and to an infection. [15]



Host Immunity

Transmission Pathways

Spreads mainly via the fecal–oral route and persists on surfaces due to environmental resilience. In hospitals, biofilm formation enables transmission through contaminated equipment and hands, while endogenous strains cause infections when host defenses weaken. [6][7][8][9][10]

Immune response

Host recognition of E. faecalis via TLR2 triggers inflammation and neutrophil recruitment, yet the bacterium’s stress-response and biofilm genes promote persistence. It resists clearance, alters cytokine signaling, and causes chronic or site-dependent immune modulation. [11][12][13][14]

Possible treatments

Preventative measures

- Proper hygiene in hospitals
- Consuming the prescribed amount of antibiotics for the prescribed duration to reduce the chances of bacterium mutating and developing resistance to the particular antibiotic
- Hosting educational campaigns to raise public knowledge.

Possible therapies

- Combining two beta-lactam antibiotics, such as cephalosporins, which work synergistically & bind with penicillin-binding proteins, inhibiting cell wall synthesis, and hence, destroying the cell. [17]
- Combining the cell wall-active agents, such as amoxicillin, with aminoglycosides such as gentamicin. [17][18]
- Ozonated Aloe Vera diluted in 70% ethanol and clorophorm showed moderate results for treating root canal infections caused by the bacteria. [19]

Conclusion

Even though the bacterium is intrinsically resistant to the individual antibiotics, it can be destroyed by combining the antibiotics together. However, due to the plasticity of the Enterococci’s genome, a definitive cure has not been found and likely will not be found which makes preventative measures a high priority. [20]



REFERENCES

THE IRON FIST OF *H. INFLUENZAE*

university of groningen

YARA VAN EERDE (S4754751), EVI GELLING (S5249813), GERLIN KNOBBE (S5161134), WOUTER OLDERODE (S5726069), MARIT RAEMAEKERS (S5371007), STEIJN VERHEIJEN (S5189098)

INTRODUCTION

Haemophilus influenzae was the first bacterial genome fully sequenced (1). It commonly lives harmlessly in the nasopharynx but can cause severe infections - especially in children and immunocompromised individuals - when it enters the bloodstream (2). The most dangerous strain, type b (Hib), can lead to sepsis, meningitis, or pneumonia (3). Fortunately, an Hib vaccine is already widely used in infants to battle the devastating consequences of the bacteria. A key factor in the pathogenicity of *H. influenzae* is its ability to acquire iron from the host, which is essential for its survival (4). The three proteins, that we studied for this research, are all involved in the infection rate of the bacteria. Understanding how these proteins function and interact can help identify strategies to disrupt iron acquisition and reduce the bacteria's ability to cause disease.

METHODS

Bioinformatics tools used:

- BLAST
- ORFfinder
- KBase
- InterPro
- AlphaFold
- TMHMM
- iTol
- Clustal Omega
- PyMol

OAPA



Figure 1: AlphaFold prediction of OpaA structure

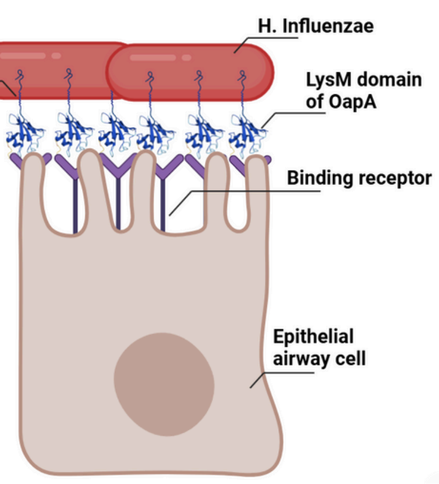


Figure 2: Schematic view of OapA mechanism with nametags. Not to scale.

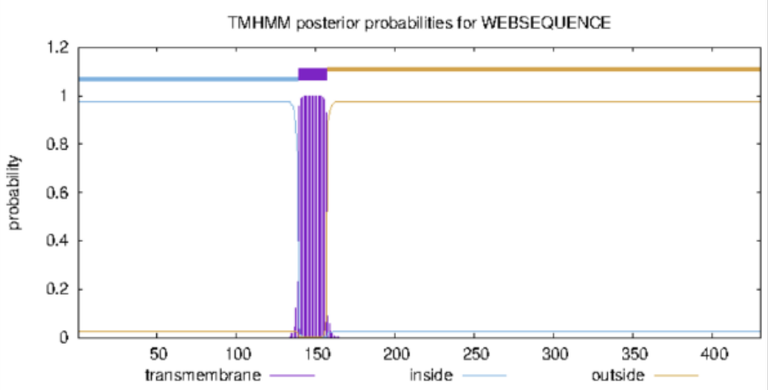


Figure 3: TMHMM result of OapA, where blue is intracellular, purple is trans-membrane and orange is extracellular

After inserting the sequence into TMHMM, OapA was found to be a trans-membrane protein. With a single TM-helix around aa-residues 140-160.

Aa residues	Domains
1-98	Intrinsically Disordered Regions (IDR)
130-160	TM Helix
351-431	LysM-like Domain

Figure 4: Interpro results of the OapA sequence put into a table. Aa residues indicates the amino acid position within the protein. "Domains" indicates the domain names.

Interpro shows an IDR at the N-terminal, which AlphaFold also shows. The first domain is around aa-residue 140-160, which is the TM-helix. The second domain is the C-terminal (LysM) domain. This C-terminal domain is also shown in AlphaFold as high confidence. The N-terminal is made up of an IDR, which can have various functions. It is believed to play a role in the phase regulation of *H. influenzae*, meaning it's useful in adapting to changing environments.

TDHA

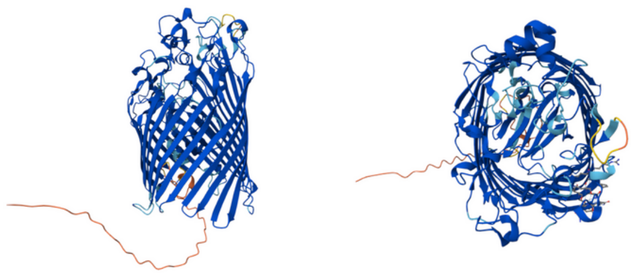


Figure 5: AlphaFold 'side' and 'top' view of TonB dependent heme receptor A

β-barrel domain: Forms the transmembrane channel through the outer membrane. It creates a cylindrical pore through which heme can be transported once the receptor is activated by the TonB system.

Extracellular loops: The loops on the top of the β-barrel face the extracellular environment. Involved in substrate recognition, binding and specifity of the receptor for heme.

Plug Domain: located inside the β-barrel. It blocks the β-barrel in the resting state, but opens through conformational change upon heme binding and TonB interaction.

N-terminal signal peptide: The signal peptide directs the protein to the outer membrane.

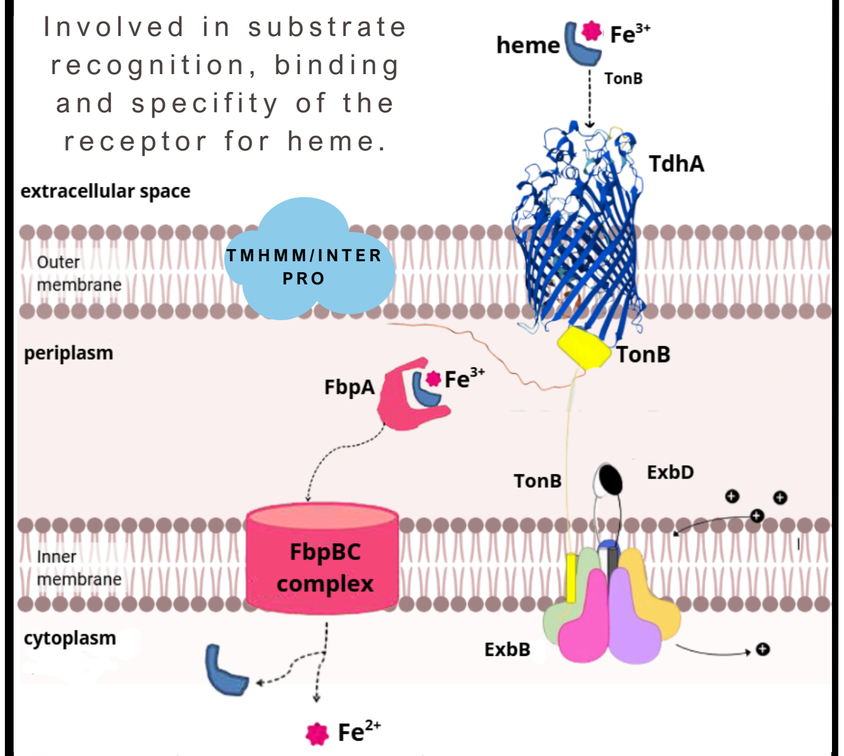


Figure 6: Schematic view of the iron transport system.

FBPC2

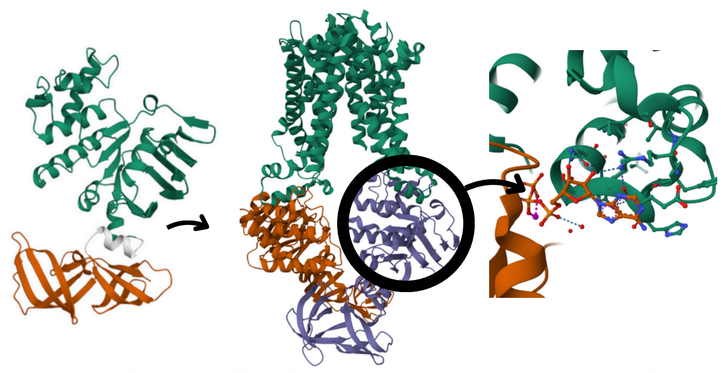


Figure 7: Protein FbpC2 visualized using Mol*. First image shows a single FbpC2 protein, the second one shows it in a complex with FbpB, and the third one shows FbpC2 with ATP bound.

After TdhA transfers Fe^{3+} to the periplasm, the FbpABC transporter takes over the ferric iron to transfer it over the inner membrane. The role of FbpC2 is the ATP hydrolysis to generate the energy needed to perform this transfer. FbpC2 is found in a complex together with FbpB, and binds ATP with residues of its alpha helices (Figure 7).

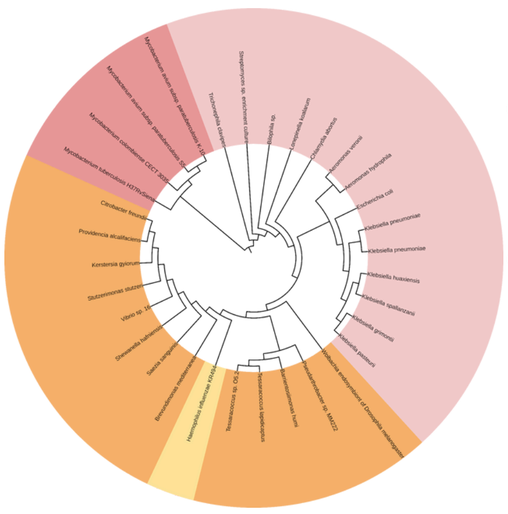


Figure 8: Phylogenetic tree of proteins similar to FbpC2 in other organisms. Closely related organisms are coloured alike and positioned in clades. *H. influenzae* is coloured in yellow. Produced using NCBI database and iTol.

Figure 8 shows the phylogenetic tree. Visible is that there are multiple organisms in which a protein is highly alike to FbpC2. Before the phylogenetic tree was made, a multiple sequence alignment was performed. This showed that the sequences in different organisms are quite alike. We thus consider the sequences to be conserved amongst organisms.

CONCLUSION

By using the mentioned bioinformatics tools, we were able to shed light on the pathogenic mechanisms of *H. influenzae* by analyzing the chosen fragments. We were able to identify three proteins of interest in the fragments; OapA, TdhA and FbpC2. These proteins all have their own function in the pathogenicity of *H. influenzae*. While OapA is important for the creation of colonies in the airway, TdhA and FbpC2 are important for the acquisition of Fe^{3+} which is necessary for the function of the bacterium.

Existing vaccines, like the Hib vaccine, target polysaccharides. Future therapies could instead focus on blocking heme acquisition (5). This approach could serve as a preventive strategy, offering new vaccine possibilities against *H. influenzae* and helping to reduce child mortality rates.

REFERENCES